Associations between Selected Dietary Factors, Selected Obesity-Related Metabolic Markers (Leptin, C-peptide, and High-sensitivity C-reactive Protein), and Lung Cancer: A Matched Case-Control Study Nested in the Prospective PLCO Trial

by

Yixian (Crystal) Chen, BSc, Brock University

In partial fulfillment of the requirements for the degree of

Master of Science in Applied Health Sciences

(Health Sciences)

Faculty of Applied Health Sciences

Brock University

St. Catharines, Ontario, Canada

ABSTRACT

The purpose of the study was to evaluate the associations between selected dietary factors, body mass index (BMI), selected obesity-related metabolic markers, and lung cancer risk as well as histological types in ever-smokers (former and current-smokers). Characteristics of interest included BMI at age 50, fruits and vegetables daily frequency, supplemental beta-carotene intake, C-peptide (CP), high-sensitivity C-reactive protein (hsCRP), and leptin concentrations. Data from the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial were analyzed. Linear regression models were used to describe the associations between quantitative variables. The relationships between variables of interest and lung cancer were studied by logistic regression modelling. Multivariable fractional polynomial (MFP) models were utilized to address non-linearity in these associations. Higher fruits and vegetables daily frequency and supplemental beta-carotene intake were associated with a lower risk of lung cancer in ever-smokers. Metabolic markers, C-peptide and hsCRP, were positively associated with lung cancer risk. Inverse relationships were observed between BMI and leptin with lung cancer risk. The relationships between selected dietary factors, BMI, selected metabolic markers, and lung cancer risk were more prominent in non-small cell lung cancer (NSCLC) in comparison with those in small cell lung cancer (SCLC).

KEY WORDS: Lung cancer, public health, diet, obesity, metabolic marker

ACKNOWLEDGEMENTS

First of all, I would like to thank all the faculty, staff, and my fellow students from the Graduate Studies, the Library, and the Department of Health Sciences at Brock University. They make the university my second home. This thesis would not have been possible without their help.

I am especially grateful to my thesis supervisor Dr. Martin Tammemägi for his constant guidance and patience throughout my master's studies. Dr. Tammemägi has been inspiring and challenging me to become a lifelong learner and a better researcher. I would like to thank my thesis committee members, Drs. Jason Liu and Joanne Crawford, for their precious advice and encouragement. Special thanks to Dr. Crawford for her insightful and constructive comments and to Dr. Tammemägi and Dr. Liu for generously devoting their time and efforts to constructing and editing this thesis. It is my privilege and honour to have been mentored by this team. A special acknowledgement goes to my external examiner, Dr. Chi-Chen Hong for her in-depth reading of this thesis and discerning commentaries on the statistical methodology and rationalizations. Furthermore, I would like to thank Dr. Terrance Wade for being the chair at my defence and for giving me the inspiration to take the path of health research since my undergraduate studies.

Special thanks are owed to my parents who have been supporting me for many years of education. Finally, I would like to express my appreciation to the participants and investigators, whose dedication makes it possible for me to use all the information needed for this study. I hope that my efforts will make a difference in their lives.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AD – Adenocarcinoma

AIC – Akaike Information Criterion

AJCC – American Joint Committee on Cancer

ALK – Anaplastic Lymphoma Kinase Gene

ATBC – Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study

BAC – Bronchiolo-Alveolar Carcinoma

BIC – Bayesian Information Criterion

BMI – Body Mass Index

CARET – Carotene and Retinol Efficacy Trial

CCSAC—Canadian Cancer Statistics Advisory Committee

CDC – Centers for Disease Control and Prevention

CI – Confidence Interval

COPD – Chronic Obstructive Pulmonary Disease

CP – C-peptide (Connecting Peptide)

CRP – C-reactive Protein

CT – Computed Tomography

DNA – Deoxyribonucleic Acid

EGFR – Epidermal Growth Factor Receptor

EML4 – Echinoderm Microtubule Associated Protein Like 4 Gene

ER – Estrogen Receptor

FFQ – Food Frequency Questionnaire

GPRP – Gastrin-Releasing Peptide Receptor

GSTM1 – Glutathione S-transferase M1 Enzyme

GWAS – Genome Wide Association Studies

HIV – Human Immunodeficiency Virus

HR – Hazard Ratio

hsCRP – High-sensitivity C-reactive Protein

IARC – International Agency for Research on Cancer

IGF-1 – Insulin-like Growth Factor-1

IQR – Interquartile Range

LC – Lung Cancer

LCC – Large Cell Carcinoma

LN – Natural Log Transformed

MeS – Metabolic Syndrome

MFP – Multivariable Fractional Polynomial

MI – Multiple Imputation

MRI – Magnetic Resonance Imaging

NCI – National Cancer Institute

NHANES III – Third National Health and Nutrition Examination Survey

NIH – National Institutes of Health

NSCLC – Non-Small Cell Lung Cancer

OR – Odds Ratio

PA – Physical Activity

PET – Positron Emission Tomography

PLCO – Prostate, Lung, Colorectal, and Ovarian

RAE – Retinol Activity Equivalents

RCT – Randomized Control Trial

RR – Relative Risk

SCC - Squamous Cell Carcinoma

SCLC – Small Cell Lung Cancer

SD – Standard Deviation

SES – Socioeconomic Status

SHS – Secondhand Smoke

T2DM – Type 2 Diabetes Mellitus

TB – Tuberculosis

TNM – Tumor-Node-Metastasis

UICC – Union for International Cancer Control

VALSG – Veterans Administration Lung Study Group

VIF – Variance Inflation Factors

WCRF/AICR – World Cancer Research Fund/American Institute for Cancer Research

WHO – World Health Organization

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

In Canada, approximately 1 in 11 males and 1 in 14 females would develop lung cancer (LC), and this type of cancer causes more cancer deaths than the total mortality from breast, colorectal, and prostate cancers together (Canadian Cancer Statistics Advisory Committee[CCSAC], 2017). Although lung cancer is an exceedingly lethal disease, there are hopes of preventing it. Risks associated with the disease may be reduced or prevented with further understanding of factors that contribute to lung cancer risk other than tobacco use.

According to multiple studies, the most significant cause of lung cancer is cigarette smoking (Alberg, Brock, Ford, Samet, & Spivack, 2013; de Groot, Wu, Carter, & Munden, 2018). About 85% of lung cancer incidence cases could be attributed by cigarette smoking (Chyou, Nomura, & Stemmermann, 1992) In Canada, 11.1% of people with lung cancer were caused by low fruits consumption and 5.3% were caused by low non-starchy vegetables consumption, respectively (Poirier et al., 2019). Higher body mass index (BMI), an indicator of body composition, is associated with numerous deteriorating health problems. Nevertheless, an inverse association between elevated BMI and lung cancer risk has been established by many studies, which implies that greater body mass protects people against lung cancer (Goodwin & Stambolic, 2015; Renehan, Tyson, Egger, Heller, & Zwahlen, 2008). Other researchers have explored the role of different lifestyle factors, for example, nutrition and diet (Dela Cruz, Tanoue, & Matthay, 2011). Dietary components, such as fruits, vegetables, and whole grains have protective effects on lung cancer. Higher levels of serum C-peptide and high-sensitivity C-reactive protein were related to a higher lung cancer risk (Goodwin et al., 2015).

The main purpose of the current study was to determine whether the effects of dietary factors on odds of lung cancer could be explained by alterations in metabolic markers in relation to sex, smoking status, and histological types of lung cancer. We intended to identify associations among diet,

BMI, metabolic markers, and the likelihood of lung cancer. This is expected to add new understanding of factors that may prevent lung cancer.

1.1 Impact of Lung Cancer

For decades, lung cancer has been one of the most common cancers in the world, and 2.09 million new occurrences were estimated in 2018 (CCSAC, 2018). With a high mortality rate, lung cancer was responsible for 1.76 million deaths (CCSAC, 2018). And it is noteworthy that lung cancer trends differ by histologic type and sociodemographic characteristics, including sex, race/ethnicity, and age. These differences represented different historical cigarette smoking rates, duration, cessation, cigarette composition, and exposure to other carcinogens (Lewis, Check, Caporaso, Travis, & Devesa, 2014). High incidence and mortality, and poor survival make lung cancer one of the high-burden diseases to societies. A better understanding of the causes of lung cancer can help to design proper protective schemes. Consequently, the burden of the disease can be minimized by carefully implementing prevention strategies (CCSAC, 2017).

1.2 Potential Mechanisms

Cigarette smoking is the single most prominent preventable risk factor, as it has been proven to have a strong direct effect on lung cancer formation (de Groot et al., 2018). The exposure to cigarette smoking may be also indirectly linked to users' BMI and metabolism, which alters people's risk of developing the disease. Excess body weight or higher BMI, a risk factor for many common (i.e. colorectal cancer) and less common cancers (i.e. gallbladder), reduces people's risk of developing lung cancer (Goodwin & Stambolic, 2015; Renehan et al., 2008; Sanikini et al., 2018). Higher C-peptide and high-sensitivity C-reactive protein (hsCRP) are metabolic markers that may reflect the carcinogenic process, but lower risk is observed in people with higher leptin level. BMI and these metabolic markers are related to diet, so the role of dietary factors in lung cancer risk should be considered along with smoking exposure. Dietary studies have found that certain dietary patterns, such as high fat, sugar, and refined

carbohydrate consumption, are associated with higher body mass. And studies on cigarette smoking and BMI found that smokers are likelier to be leaner than non-smokers. On the other hand, evidence from observational studies have substantiated the role of a healthy diet in lung cancer. Fruits and vegetables consumption is inversely related to the incidence of the disease (Alberg et al., 2013). In addition, an earlier study of the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial found that several obesity-related metabolic markers are associated with lung cancer risk (Goodwin et al., 2015). Potential interaction of associations with sex were examined since metabolism differs by biological sex. Therefore, diet might have an impact on the likelihood of lung cancer through regulation of BMI, leptin, C-peptide, and hsCRP; and the effects might differ by sex, and smoking status.

1.3 Literature Gaps

Previous studies established the relationship between smoking and metabolic changes, including lower leptin, higher C-peptide, and hsCRP (Cibickova et al., 2015; Helmersson, Larsson, Vessby, & Basu, 2005; Perkins & Fonte, 2002). Although the results of dietary studies were inconsistent, a well-accepted finding suggested that a dietary pattern with high intakes of fruits, vegetables, and whole grains was inversely associated with leptin and C-reactive protein (CRP) levels independent of BMI (Ko et al., 2016; Nettleton et al., 2006). Higher fruits or green leafy vegetables intake was associated with lower risk of type 2 diabetes mellitus (T2DM) (Li, Fan, Zhang, Hou, & Tang, 2014). This might be explained by the antioxidant content in fruits and vegetables that reduces systemic oxidative stress and inflammatory responses. Carotenoid metabolism changes during inflammation, and carotenoid level can also influence inflammatory responses and oxidative stress (Rubin, Ross, Stephensen, Bohn, & Tanumihardjo, 2017). Hence, beta-carotene may play a role in the development of lung cancer by mediating inflammation. hsCRP belongs to the family of innate immune response proteins and is considered a marker of inflammatory processes (Bassuk, Rifai, & Ridker, 2004). Furthermore, the associations between pre-diagnostic metabolic markers and lung cancer risk regarding smoking status

were investigated. Higher C-peptide and hsCRP concentrations were associated with higher lung cancer risk, while leptin level and BMI were inversely related to the risk in ever-smokers (former and current-smokers) (Goodwin et al., 2015). A moderate positive correlation between BMI and leptin level suggested a possible association, but a causal relationship could not be determined yet (Altman & Krzywinski, 2015; Kazmi et al., 2013). Moreover, BMI was documented to be inversely associated with lung cancer risk from a previous study using data from the PLCO Cancer Screening Trial (Tammemagi et al., 2011).

The study designs of previous studies were mostly cross-sectional or retrospective case-control, so they were possibly compromised by reverse causality. For instance, the metabolic changes were caused by lung cancers. The current case-control study evaluated the effects of dietary factors, BMI at age 50, leptin, C-peptide, and hsCRP on lung cancer risk. It was nested in the experimental cohort of the PLCO Cancer Screening Trial. A nested case-control design possesses a higher position in the hierarchy of study validity and brings more evidence for establishing causality between exposures and the outcome of interest. In addition, the interactions of participants' sex, and smoking status had not been completely examined yet; the current study evaluated the associations between dietary factors, BMI at age 50, metabolic markers, and lung cancer adjusted for potential confounders.

1.4 Response to Gaps in Current Knowledge

To address the gaps in knowledge, the current research was conducted using the pre-existing data from the PLCO Cancer Screening Trial in the United States. The PLCO Trial was a randomized prospective study launched by the National Cancer Institute (NCI) in 1993. The metabolic markers and dietary data were collected before the patients were diagnosed with lung cancer. Hence, the study could appropriately clarify the temporal aspects of the associations and might suggest possible causal relationships. Moreover, if the identified mechanisms of associations in the study can be validated in

future studies, this may lead to a discovery of potential sites for intervention. Consequently, the results may provide future research directions.

1.5 Study Aim, Objectives and Hypotheses

Study Aim: To evaluate the associations between selected dietary factors, BMI, metabolic markers (leptin, C-peptide, and high-sensitivity C-reactive protein), with each other and with lung cancer.

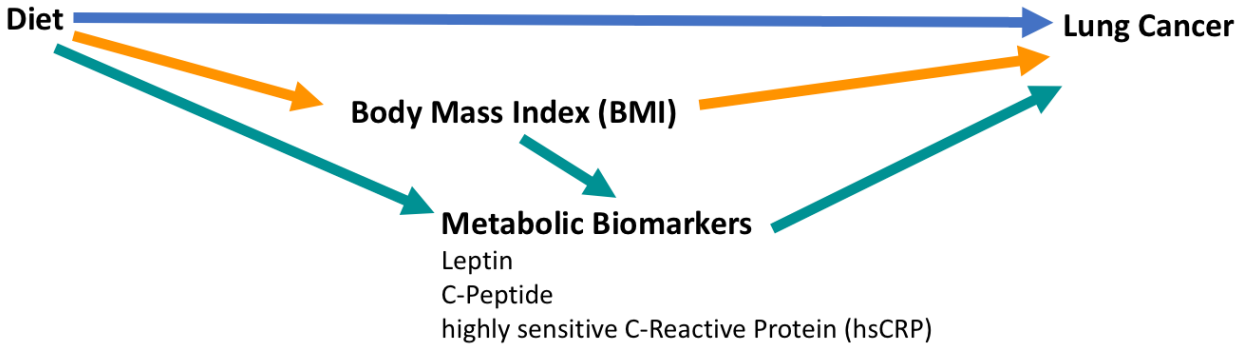

Figure 1. Schematic Presentation of Potential Pathways

Study Objectives:

(1a) Describe the associations between selected dietary factors (fruits and vegetables daily frequency and supplemental beta-carotene intake) and lung cancer, and (1b) if the associations differ by potential effect modifiers (sex, smoking status, and histological types of lung cancer).

2a) Describe the associations in the pathway between Diet-BMI-LC, (2b) evaluate if the associations differ by potential effect modifiers, and (2c) and if BMI in part explain the association between diet and lung cancer.

(3a) Describe the associations in the pathway between Diet-Metabolic markers-LC, (3b) evaluate if the associations differ by potential effect modifiers, (3c) and if metabolic markers in part explains the associations between diet and lung cancer, and (3d) between BMI and lung cancer.

Hypotheses:

Relationship between diet and lung cancer:

H_0: Dietary factors are not associated with lung cancer as estimated by odds ratio approaching the null value.

H_1: Dietary factors are associated with lung cancer as estimated by odds ratio away from the null value.

Relationship between diet and BMI:

H_0: Dietary factors are not associated with BMI as estimated by slope approaching the null value.

H_1: Dietary factors are associated with BMI as estimated by slope away from the null value.

Relationship between BMI and lung cancer:

H_0: BMI is not associated with lung cancer as estimated by odds ratio approaching the null value.

H_1: BMI is associated with lung cancer as estimated by odds ratio away from the null value.

Relationship between diet and metabolic marker level:

H_0: Dietary factors are not associated with metabolic markers as estimated by slope approaching the null value.

H_1: Dietary factors are associated with metabolic markers as estimated by slope away from the null value.

Relationship between metabolic markers level and lung cancer:

H_0: Metabolic markers are not associated with lung cancer as estimated by odds ratio approaching the null value.

H_1: Metabolic markers are associated with lung cancer as estimated by odds ratio away from the null value.

The main dependent variable a person's lung cancer status (yes/no) is dichotomous, and the independent explanatory variables are quantitative (e.g. supplemental beta-carotene intake) and categorical variables (e.g. education). Cases and controls were matched on sex, age, race/ethnicity, smoking status, year of study entry, duration of follow-up, and study center. Therefore, conditional logistic regressions examined the associations for matched data, the method estimated the effects of

multiple explanatory variables with and without adjustments for relevant variables and identified effect modifications (also known as interactions). Hypotheses of interactions and correlations between metabolic markers are not stated here because those evaluations were of an exploratory nature.

1.6 Conclusion

The pooled five-year survival rate of lung cancer patients for all stages was just 17% (CCSAC, 2019). And the deadly disease is not uncommon, approximately 8.7% of males and 7.1% of females in Canada would develop the disease (CCSAC, 2017). Patients and their families suffer as do communities worldwide. Gaining further understanding of additional risk factors for lung cancer might lead to preventive measures, which would be the most cost-effective strategy for individuals and societies to manage the harms of the disease (Tota, Ramanakumar, & Franco, 2014). The current study aimed to contribute to the existing body of knowledge by evaluating the associations among diet, BMI, metabolic markers, and lung cancer. The associations described in the study might help to clarify the inverse relationship between pre-diagnostic BMI and lung cancer. The mechanism of dietary factors elucidated in the current study may offer new targets for interventions for primary prevention. A better understanding of selected dietary factors, BMI, and metabolic markers may improve a screening program and a prediction tool for lung cancer. The additional components of metabolic markers provided by the study may further improve the performance of risk assessment tools and help to identify individuals at high risk for developing lung cancer.

CHAPTER 2 REVIEW OF LITERATURE

2.1 Overview and Lung Cancer Statistics

Lung cancer is the most common cancer and the most common cause of cancer death in the world. The crude and age standardized incidence and mortality rates in 2018 were 22.5 and 18.6 per 100,000, which contributed to 2.09 million new cases and 1.76 million deaths (International Agency for Research on Cancer [IARC], 2018). The disease continues to be the most common cancer in males worldwide. The incidence rate for females is lower than males because of different tobacco use. Since the overall ratio of mortality to incidence was 0.87, it is a highly fatal disease (IARC, 2018).

In Canada, about 8.7% males and 7.1% females would develop lung cancer in their lifetimes (CCSAC, 2017). It is a leading cause of cancer death as well, killing more Canadians than the other three major cancer types combined (breast, colorectal, and prostate cancers). Overall, lung cancer accounted for approximately 14% of all new cases in both males and females (CCSAC, 2017). In Canada, it was projected that 28,600 people would develop lung cancer in 2017 and 21,100 would die from it (Government of Canada, 2017). The age standardized incidence remained higher in males (76.5 per 100,000) than females (65.3 per 100,000) (Government of Canada, 2017). The disparities in incidence rates between sexes is explained by historical differences in tobacco use. In males, a noticeable decline in the prevalence of daily smokers occurred in the mid-1960s in Canada, followed by a decrease in incidence of lung cancer about 20 years later. A downturn in smoking did not emerge among females until the mid-1980s, which indicates that incidence rate of lung cancer in females may start to drop in the coming years (CCSAC, 2017). In contrast to many other cancers where mortality increases with age, deaths from lung cancer remain the highest in one age range: 70 – 79 years old for both sexes (CCSAC, 2017). Five-year relative survival rate for lung cancer was about 17%, with 14% in males and 20% in females (CCSAC, 2017). This means that comparing to 100 people who are alive at the end of five-year period in the general population, the number of people with lung cancer, similar age and sex who will be

alive five years after diagnosis is only 17. Moreover, relative survival rate decreases as patients' ages go up. People aged 15 – 39 years at diagnosis have the highest five-year relative survival rate of 45%, while individuals aged 80 – 99 years have the lowest rate of 10% (CCSAC, 2017). It can be explained that older individuals with lung cancer are more likely to die from competing causes of death than younger individuals.

2.2 Biology of Lung Cancer

2.2.1 Etiology and Pathogenesis

Normal cells grow and divide to produce new cells when the body requires them, and newly generated cells take the place of aged and damaged ones. When the balance is disrupted, aged or damaged cells persist in living and unrequested new cells are produced disproportionately (National Cancer Institute [NCI], 2015b). As a result, these extra cells replicate interminably and may progress and form tumors. Particularly, a malignant tumor, also known as cancer, invades tissues nearby and spreads to distal sites of one's body (NCI, 2015b). Thereby, having a sound grasp of irregular cell activities is vital to understanding carcinogenesis. One cause of abnormal cell activity is genetic alterations. The genetic changes can occur through a variety of mechanisms, including inherited from one's biological parents, acquired during his/her lifetime by reason of mutations, or triggered by certain environmental exposures (NCI, 2015b). In general, the alterations can take place in 3 types of genes: proto-oncogene, tumor suppressor gene (also known as antioncogene), and DNA repair gene. Proto-oncogenes are obligated to maintain normal development and reproduction of cells. So, changes in these genes may result in hyperactivity, and they may turn into cancer-causing oncogenes. Similarly, mutation in antioncogenes may result in loss of function which leads to unrestrained cell growth. By contrast, DNA repair genes have a capacity to refurbish injured genes, so cells with genetic alterations in them are more inclined to accumulate additional mutations in other genes. Because of this, cells that have undergone multiple mutations can eventually become cancerous (NCI, 2015b). In other words, cancer is

a joint consequence of higher exposures to detrimental factors and higher susceptibility to these factors among people (Alberg et al., 2013). Moreover, the interaction between two harmful exposures can act synergistically and increase lung cancer risk that is considerably greater than the sum of two individual impacts. For example, exposure to asbestos and radon gas are well-recognized risk agents which can amplify the harmful effect of cigarette smoking (Saracci, 1987).

2.2.2 Classification and Histology

Based on the cell type of a malignant tumor, lung cancer can be categorized into two general histological classes: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). The former type accounts for 80 to 85 percent of all cases, and it is comprised of adenocarcinoma (AD), squamous-cell carcinoma, and large cell carcinoma (American Cancer Society, 2016b; Collins, Haines, Perkel, & Enck, 2007; Hoffman, Mauer, & Vokes, 2000).

Small Cell Lung Cancers (SCLCs)

Small cell lung cancers (SCLCs) account for about 14% of all lung cancer cases, and unique clinical and histological characteristics distinguish them from NSCLCs. SCLCs are made of small cells with little cytoplasm, poorly marked cell borders, finely granular nuclear chromatin, and inconspicuous nucleoli (Travis, Brambilla, Muller-Hermelink, & Harris, 2004). SCLCs are rarely seen in non-smokers, and are likelier to progress considerably, and metastasize readily in comparison to NSCLC (CCSAC, 2017). The following will discuss cell types specific to NSCLC.

Non-small cell lung cancer

Adenocarcinomas (ADs)

Adenocarcinomas account for roughly 40% of all lung cancers (Collins et al., 2007; Travis, 2011). Although adenocarcinomas manifest primarily in former and current-smokers, they are the most prevalent type found in non-smokers (CCSAC, 2018). Given that lung cancer occurs less frequently in non-smokers, the number of adenocarcinoma cases are lower in this population compared to that of

smokers (CCSAC, 2018). Moreover, adenocarcinomas occur more often in females than males, and have become a common type in young never-smokers (American Cancer Society, 2016b; Dela Cruz et al., 2011).

Adenocarcinomas are closely related to the popularity of filter-tipped cigarettes. The tobacco industry adds filter tips to cigarettes to cut down tar and nicotine yields (Ernst L Wynder & Muscat, 1995). In order to compensate for the lower delivery of nicotine, users tend to inhale more deeply and smoke more intensely than non-filter smokers. A greater draw of smoke can counteract the decrease in tar exposure (Ernst L Wynder & Muscat, 1995). The peripheral bronchioles are defective in defense mechanisms. In consequence, the distant lung tissue is exposed to higher dose of carcinogens, making people more susceptible to developing adenocarcinomas. In other words, more adenocarcinomas were detected when filtered cigarettes began to dominate the market.

Squamous-Cell Carcinomas (SCCs)

About 25 to 30% of all lung cancers are squamous-cell carcinomas, and they are characterized by thin, flattened cells with keratinized cytoplasm and intercellular bridges (American Cancer Society, 2016b; Travis, 2011). SCCs are typically linked with a history of smoking and generally occur in the central part of the thorax, especially in the vicinity of a major airway or a bronchus (American Cancer Society, 2016b). In addition, SCCs is a predominant subgroup in male smokers (Khuder, Dayal, Mutgi, Willey, & Dayal, 1998; Pesch et al., 2012).

Large Cell Carcinomas (LCCs)

Approximately 10 to 15% of lung cancers belong to LCCs. They can be found in any region of the lungs. Because LCCs frequently progress and spread quickly, healthcare providers have difficulty in managing and controlling them (American Cancer Society, 2016b).

2.3 Tumor Staging and Grade

2.3.1 Cancer Staging

Cancer staging represents the severity of a patient's cancer referring to the size of a primary tumor and if it has metastasized to other sites of his/her body. Staging enables doctors to develop an appropriate treatment plan and make an accurate prognosis for the patient (American Joint Committee on Cancer [AJCC], 2018; NCI, 2015a). Regardless of different types of staging, all systems consist of four principal elements: (1) the location of a primary tumor in the body; (2) the magnitude of the tumor and extension of tumor locally; (3) if the tumor has spread to adjacent lymph nodes; and (4) whether the cancer has metastasized to a distant location of the body. The Tumor-Node-Metastasis (TNM) classification system (8th edition), developed and maintained by the American Joint Committee on Cancer (AJCC, 2018) and the Union for International Cancer Control (UICC), is the most widely utilized tool by medical professionals. As shown in Table 1, there are three anatomical components in the system: T for the size of a primary tumor in long axis, or direct extent of the tumor into contiguous structures; N for the involvement of regional lymph nodes; and M for the presence of distant metastases beyond regional lymph nodes (Detterbeck, Boffa, Kim, & Tanoue, 2017; Kay et al., 2017; Mirsadraee, Oswal, Alizadeh, Caulo, & van Beek, 2012). Moreover, each T, N, and M component is split into multiple secondary categories based on particular characteristics. In Table 2, three elements are integrated and reclassified into prognostic stage groups (Detterbeck et al., 2017; Edge, 2017; Kay et al., 2017).

Table 1. Tumor-Node-Metastasis (TNM) Cancer Staging System for Lung Cancer, 8th edition. (American Joint Commission on Cancer, 2017)

Primary Tumor (T)	
T0	No evidence of primary tumor
Tis	Carcinoma *in situ*
T1	Tumor 3 cm or less in greatest diameter
T2	Tumor more than 3 cm but no more than 5 cm or tumor involving visceral pleura, main bronchus, atelectasis to hilum
T3	Tumor more than 5 cm but no more than 7 cm or invading chest wall, pericardium, phrenic nerve or separate tumor nodule(s) in the same lobe
T4	Tumor greater than 7 cm or invading mediastinum, diaphragm, heart, great vessels, recurrent laryngeal nerve, carina, trachea, esophagus, spine; or tumor nodule(s) in a different ipsilateral lobe
Reginal Lymph Nodes (N)	
N0	No regional lymph node metastasis
N1	Metastasis in ipsilateral pulmonary or hilar nodes
N2	Metastasis in ipsilateral mediastinal/subcarinal nodes
N3	Metastasis in contralateral mediastinal/hilar, or supraclavicular nodes
Distant Metastasis (M)	
M0	No distant metastasis
M1a	Malignant pleural/pericardial effusion or pleural/pericardial nodules or separate tumor nodule(s) in a contralateral lobe
M1b	Single extrathoracic metastasis
M1b	Multiple extrathoracic metastases (1 or more than 1 organ)

Note: the table is adapted from The Eighth Edition Lung Cancer Stage Classification (2017) (Detterbeck et al., 2017)

2.3.2 Stage Groups

Cancers at Stage I are usually small or have not grown deeply into nearby healthy tissues. Stage II and III indicate that cancers have grown in size and may have invaded to lymph nodes or adjacent tissues as well. When the tumor has spread to other organs or sites of the body, the individual is diagnosed with Stage IV cancer (Edge, 2017).

Table 2. Stage Grouping for Lung Cancer, 8th edition. (American Joint Commission on Cancer, 2017)

T/M	Description	N0	N1	N2	N3
T1	Tumor ≤ 1 cm	IA1	IIB	IIIA	IIIB
	Tumor > 1 but ≤ 2 cm	IA2	IIB	IIIA	IIIB
	Tumor > 2 but ≤ 3 cm	IA3	IIB	IIIA	IIIB
T2	Tumor < 3 but ≤ 5 cm or tumor involving: visceral pleura[a], main bronchus (not carina), atelectasis to hilum[a]	IB	IIB	IIIA	IIIB
	Tumor > 3 but ≤ 4 cm	IB	IIB	IIIA	IIIB
	Tumor > 4 but ≤ 5 cm	IIA	IIB	IIIA	IIIB
T3	Tumor > 5 but ≤ 7 cm	IIB	IIIA	IIIB	IIIC
	or invading chest wall, pericardium, phrenic nerve	IIB	IIIA	IIIB	IIIC
	or separate tumor nodule(s) in the same lobe	IIB	IIIA	IIIB	IIIC
T4	Tumor > 7 cm	IIIA	IIIA	IIIB	IIIC
	or tumor invading: mediastinum, diaphragm, heart, great vessels, recurrent laryngeal nerve, carina, trachea, esophagus, spine;	IIIA	IIIA	IIIB	IIIC
	or tumor nodule(s) in a different ipsilateral lobe	IIIA	IIIA	IIIB	IIIC
M1	Malignant pleural/pericardial effusion[b] or pleural/pericardial nodules	IVA	IVA	IVA	IVA
	or separate tumor nodule(s) in a contralateral lobe	IVA	IVA	IVA	IVA
	Single extrathoracic metastasis	IVA	IVA	IVA	IVA
	Multiple extrathoracic metastases (1 or > 1 organ)	IVB	IVB	IVB	IVB

Note: the table is reprinted from The 8th Edition Lung Cancer Stage Classification (2017) (Detterbeck et al., 2017)
cm: centimeter
[a] such tumors are classified as T2a if >3 and ≤4 cm, T2b if >4 and ≤5 cm
[b] Pleural effusions are excluded that are cytologically negative, non-bloody, transudative, and clinically judged not to be a cancer

2.3.3 Tumor Grade

In contrast to staging, tumor grade represents the irregularity and deviation of a tumor from healthy tissues, it is determined by the appearance of the tumor tissue under a microscope and its growth rate (NCI, 2013). Shown in Table 3, tumors made of cells and tissues resemble normal ones are described as "well-differentiated" (low grade); they have a lower level of aggressiveness and tend to spread slowly (NCI, 2013). In contrast, "poorly-differentiated" or "undifferentiated" (high grade) cells signifies the presence of aggressive tumors that have lost the normal structure of cells and they grow

and spread rapidly. Patients with high grade tumors usually have poor prognoses and are in need of immediate therapy (NCI, 2013).

Table 3. Tumor Grading System

Grade	Description
GX	Grade cannot be assessed (undetermined grade)
G1	Well differentiated (low grade)
G2	Moderately differentiated (intermediate grade)
G3	Poorly differentiated (high grade)
G4	Undifferentiated (high grade)

Note: the table is adapted from *AJCC Cancer Staging Manual*. 8th edition (2017) (Edge, 2017)

2.4 Treatment and Prognosis

Surgical resections, especially a minimally invasive procedure like video-assisted thoracic surgery, followed by chemotherapy are recommended for cases with removable stage I and II NSCLC, (Collins et al., 2007; Howington, Blum, Chang, Balekian, & Murthy, 2013; Latimer & Mott, 2015). The five-year survival rate is 60 to 70% for stage I and 40 to 50% for stage II patients (Collins et al., 2007). Based on performance status and lung cancer histology, chemotherapy and radiotherapy are recommended in stage III patients for the purpose of eliminating visible intrathoracic tumors and preventing subsequent extrathoracic metastases (Latimer & Mott, 2015; Ramnath et al., 2013; West & Jin, 2015). Five-year survival rate is between 15 and 30% for surgical removable stage IIIA, 10 to 20% for irremovable and advanced stage IIIA and stage IIIB with the involvement of contralateral or supraclavicular lymph nodes (Collins et al., 2007). For stage IIIB (pleural effusion) or stage IV cases, two-year survival rate is between 10 to 20%. For patients with stage IV tumor, the main purpose of treatment is to improve quality of life (Latimer & Mott, 2015; Temel et al., 2010). Palliative care is aimed at managing symptoms, providing psychological support and assisting with decision-making for terminally ill patients and their families. It should be introduced early for patients with metastatic NSCLC or at any stage when their morbidity is severe (Latimer & Mott, 2015; Temel et al., 2010).

Along with the TNM classification system, the Veterans Administration Lung Study Group (VALSG) two-stage classification scheme clinically categorizes SCLCs into limited-stage and severe

extensive-stage (Jett, Schild, Kesler, & Kalemkerian, 2013). For people with limited-stage SCLCs, surgery may be recommended. For either staged SCLCs, chemotherapy and radiation therapy with a platinum-based agent in conjunction with either etoposide or irinotecan may be suggested at the same time (Jett et al., 2013). In addition, opportune palliative care is recommended concerning the severity of cancer, morbidity, and patients' preferences (Latimer & Mott, 2015). People with limited-stage SCLCs are treated with the intention of curing the disease; however because of SCLC's aggressive nature, the five-year survival rate is between 20 to 25%. The figure is virtually zero among extensive-stage patients (Jett et al., 2013; Latimer & Mott, 2015).

The survival rates of lung cancer are low, especially for people diagnosed at later stages (CCSAC, 2018). For instance, estimates of five-year relative survival rates for NSCLCs range from 68% to 92% for stage I, but the number is between 1% to 10% for stage IV (American Cancer Society, 2017). Moreover, five-year survival estimates are pertinent to histologic types, for example 31% for SCLCs stage I and greater than 68% for NSCLCs stage I (American Cancer Society, 2016a; American Cancer Society, 2017).

About 50% of all lung cancers patients are identified at stage IV (CCSAC, 2018). Nearly one-quarter (23.1%) of NSCLCs are identified at stage I compared with only 3.5% of SCLCs. Because most cases are diagnosed at advanced stages and tumors have already spread, lung cancer has one of the lowest relative survival rates of all the major types of cancer in Canada (CCSAC, 2018).

2.5 Clinical Presentation and Diagnosis

2.5.1 Signs and Symptoms

Although approximately 10% of lung cancer cases show no symptoms, most patients manifest some signs or symptoms at the time of diagnosis (Beckles, Spiro, Colice, & Rudd, 2003; Collins et al., 2007; Latimer & Mott, 2015). Some symptoms affect individuals' whole bodies non-specifically, such as fatigue, loss of appetite, and weight loss. And other cases present with specific symptoms caused by a primary or a secondary tumor (Collins et al., 2007). A primary tumor often induces symptoms like

coughing, coughing up of blood, labored breathing, pain or discomfort of the chest (Collins et al., 2007; Latimer & Mott, 2015). Intrathoracic spread caused by direct invasion of a tumor or lymphangitic metastasis appears in 40% of cases, and common presentations included chest wall invasion, esophageal symptoms, Horner's syndrome, Pancoast's tumor, phrenic nerve paralysis, pleural effusion, recurrent laryngeal nerve paralysis, and superior vena cava obstruction (Collins et al., 2007). And lung cancer patients may experience unconventional symptoms when secondary tumors affect their bodies. For example, a patient may experience difficulty in swallowing when the esophagus is affected by a secondary intrathoracic tumor. And a Pancoast tumor, in rare cases of NSCLCs, is atypically found at the top end (apex) of either lung. Common symptoms associated with lung cancer are seldom found in these patients. Instead, a Pancoast tumor is often accompanied by pain in the shoulder and arm (Hamilton, Peters, Round, & Sharp, 2005). Nearly one-third of patients display signs of extrathoracic spread, and their symptoms are related to the location of distant metastases (Collins et al., 2007; Latimer & Mott, 2015). A paraneoplastic syndrome is another consequence of the tumor or immune response through bioactive substances secretion. The syndrome occurs in about 10% of lung cancer cases, and common examples at presentation are digital clubbing, hypercalcemia, and hyponatremia (Collins et al., 2007). In addition, two symptoms, hemoptysis and digital clubbing, are associated with a higher risk of having lung cancer (Latimer & Mott, 2015).

2.5.2 Diagnosis

For suspected lung cancer patients, health history and physical examination are evaluated at first to gather subjective data on symptoms and potential risk factors. Complete blood count and blood chemistry tests (including alkaline phosphatase or calcium, and liver function tests) provide information on overall health and possible metastatic growth in the patient. Several radiological exams are employed, for example: chest X-ray (chest radiograph), computed tomography (CT) scan, positron emission tomography (PET) scan, magnetic resonance imaging (MRI), and ultrasound (CCSAC, 2018;

Latimer & Mott, 2015). A negative result of a chest X-ray cannot completely rule out lung cancer, CT or PET are necessary when a suspicion remains high (Latimer & Mott, 2015).

Physicians diagnose the type of lung cancer based on clinical presentations and radiographic exams. Although, obtaining a tissue biopsy is essential for an accurate diagnosis and planning an appropriate treatment plan. Diagnostic assessment will be further determined based on the type of procedure, dimensions and location of the tumor, comorbidities, and potential presence of metastasis (Herth, Eberhardt, Vilmann, Krasnik, & Ernst, 2006; Lamprecht, Porsch, Pirich, & Studnicka, 2009; Latimer & Mott, 2015; Rivera, Mehta, & Wahidi, 2013). Conventional bronchoscopy works best for central lesions, whereas CT-guided transthoracic needle aspiration is typically used for peripheral lesions (Rivera et al., 2013). Endobronchial ultrasound is suitable for nodules between the lungs, and electromagnetic navigation bronchoscopy improves the diagnostic yield of bronchoscopy for peripheral lesions (Herth et al., 2006; Lamprecht et al., 2009).

2.6 Risk Factors

Risk factors of a health condition are defined as characteristics, attributes or exposures that increase the probability of developing a disease (CCSAC, 2019). A risk factor may not necessarily be a cause of the disease. But understanding risk factors of lung cancer is crucial in public health as it helps to develop preventive programs. To better understand risk factors, these characteristics are mainly classified as non-modifiable or modifiable risk factors according to whether an alteration is achievable.

2.6.1 Non-Modifiable Risk Factors

2.6.1.1 Age

The inevitable risk factor of ageing is believed to have the most pronounced effect on never-smokers, although it plays a pivotal role regardless of a person's smoking status (McCarthy, Meza, Jeon, & Moolgavkar, 2012; Thun et al., 2006). Meanwhile, people are living longer lives with the average age of most populations around the world climbing, thus lung cancers cases will increase more so than past

decades. Deoxyribonucleic acid (DNA) damage and shortening telomeres are more commonly found when people are in their advanced ages, and they are associated with lung cancer development (de Groot et al., 2018). Around 53% of patients are diagnosed with lung cancer at age 55 to 74, and 37% are over 75 years old with the median age at diagnosis of 70 for both sex groups (de Groot et al., 2018; Torre et al., 2015). Lung cancer has been found to be the leading cause of death in males over 40 years of age and in females above 59 in the United States (de Groot et al., 2018; Siegel, Miller, & Jemal, 2018).

2.6.1.2 Sex

The trends of lung cancer are changing discordantly between males and females. Historically, more males used tobacco products, and incidence and mortality rates were higher in males than in females (Siegel et al., 2018). Then, female rate of smoking began after the World War II with cessation rates lagging behind males. A rapid drop in incidence rate was found in males, but the figures in females declined slowly (de Groot et al., 2018). Even though tobacco use is the major determinant of lung cancer in females, roughly 20% of cases have never smoked. Other factors, like environmental exposures, genetic mutation, hormonal effects, and infections, are potentially involved in the development of lung cancer among female non-smokers (Kligerman & White, 2011).

The histological composition and genetic predisposition of lung cancer also differ by sex. Adenocarcinoma is the most frequently found lung cancer type in both males and females (Sagerup, Småstuen, Johannesen, Helland, & Brustugun, 2011). The proportion of females identified as non-smoking lung cancer cases is higher than that of males, and the incidence of bronchiolo-alveolar carcinoma (BAC), a subtype of adenocarcinoma, is two to four times more commonly seen in females. The probability of epidermal growth factor receptor (EGFR) mutation is greater in tumors of female patients with lung cancer (O'Keeffe & Patel, 2008; Planchard, Loriot, Goubar, Commo, & Soria, 2009). Certain DNA sequences more frequently found in female smokers probably make them more susceptible

to the disease. For example, overexpression of X-linked gastrin-releasing peptide receptor (GPRP) is associated with bronchial cell proliferation (Kligerman & White, 2011; North & Christiani, 2013).

Overall, the findings of hormonal effects cannot draw a coherent conclusion. The estrogen receptor (ER) α, absent in healthy lung tissue, has been found overexpressed in adenocarcinoma of female patients, but some studies also detected the overexpression in male cases (Kligerman & White, 2011). Randomized controlled trials (RCTs) of hormonal therapy with an estrogen and progestin formulation have ascertained a significantly higher lung cancer risk, which signals a potential hormonal influence on the etiology of lung cancer in females (Alberg et al., 2013; Greiser, Greiser, & Dören, 2010). Nonetheless, a meta-analysis of 14 cohort studies concluded that the use of previous hormonal replacement therapy had an insignificant effect on the risk of lung cancer in females (Bae & Kim, 2015).

2.6.1.3 Race/Ethnicity

Race classification used in some biomedical studies adds meaningfulness to the geographical area of origin for a person's ancestry, which contains five main categories: Whites, Blacks or African Americans, Asians, Native Hawaiians or other Pacific Islanders, and American Indians or Alaska Natives (Schabath, Cress, & Muñoz-Antonia, 2016). Race and ethnicity does not only represent collective social and cultural structures, but they are also linked to socioeconomic status (SES) (Dela Cruz et al., 2011; Schabath et al., 2016). Blacks or African Americans have the highest incidence and mortality rates of the five groups (Ryerson et al., 2016). A review from the Centers for Disease Control and Prevention (CDC, 2010) reported that the annual number of new cases of lung cancer was the highest among Blacks or African Americans (76.1 per 100,000), followed by Whites (69.7 per 100,000), American Indians or Alaska Natives (48.4 per 100,000), and Asians or Pacific Islanders (38.4 per 100,000) for the period 1998 to 2006.

A disproportion of race/ethnicity composition can also be found between sexes in relation to lung cancer incidence. For instance, Blacks or African Americans had the highest incidence rate (104.5

per 100,000) among males, and Whites had the highest incidence rate (57.6) among females (Underwood et al., 2012). Moreover, a study on trends in lung cancer incidence and mortality among 20 to 39 years old recognized a narrowing of the racial gap, and a change was expected to proceed in the following decades as a result of the substantial drop in smoking prevalence among Black or African American youth since the 1970s (Jemal, Center, & Ward, 2009). Asian ethnicity is a beneficial prognostic factor for survival independent of smoking status, and Asian patients with NSCLC respond better to chemotherapy than Whites (Alberg et al., 2013; Soo et al., 2011). A convincing explanation has not been reached so far, but the differences in tumor characteristics might contribute to some extent: the prevalence of epidermal growth factor receptor (EGFR) mutation in Asians is higher than in Whites, and patients with this genetic mutation usually react well to treatment with EGFR inhibitors, such as Gefitinib (Alberg et al., 2013; W. Zhou & Christiani, 2011).

2.6.1.4 Family Aggregation and Genetic Susceptibility

Family aggregation is the occurrence of multiple cases of lung cancer within a given family, which happens due to environmental exposures and genetic susceptibility. A large international case-control study found that family history of lung cancer in any first-degree relative (parents, full-siblings, and children) significantly increased a person's risk (Lissowska et al., 2010). More specifically, a positive family history was associated with 1.72-fold increased odds ratio for lung cancer (OR = 1.72; 95% CI, 1.56-1.88), and having two or more lung cancer cases in the family increased the odds ratio by 3.6 times (OR = 3.6; 95% CI, 1.56-8.31) (de Groot et al., 2018; Lissowska et al., 2010). A case-control study that examined never-smokers in Ontario found an insignificant association between positive family history of any cancer and odds of lung cancer, but having a family member with young onset of any cancer (age at onset < 50 years old) is marginally associated with an increased odds ratio of 1.8 (95% CI, 1.0-3.2) among never-smokers (Brenner et al., 2010).

Substantial research has been done to investigate the effect of certain genetic factors on lung cancer risk. The associations between chromosome regions 5p15, 15q25-26, 6q21, and elevated lung cancer risk have been determined by genome wide association studies (GWAS) (de Groot et al., 2018; Herbst, Heymach, & Lippman, 2008; Schwartz & Cote, 2016). The 5p15 region, acting on cell replication, is associated with a higher risk of adenocarcinomas after adjusting for smoking status (Landi et al., 2009; Lips et al., 2010). And variants at chromosome locus 6p21 increase the risk of lung cancer among never-smokers (Yokota, Shiraishi, & Kohno, 2010).

2.6.2 Modifiable Risk Factors

2.6.2.1 Tobacco Use and Secondhand Smoke

The vast majority of lung carcinomas are caused by one risk factor: the use of tobacco products, which contributes to approximately 90% of lung cancer cases in countries where smoking is common (Alberg et al., 2013; de Groot et al., 2018). Lung cancer would continue to be a common cause of death before a substantial decline in smoking prevalence is accomplished (Alberg, Brock, & Samet, 2005; Molina, Yang, Cassivi, Schild, & Adjei, 2008). In the 1930s, Müller published one of the very first case-control studies and proposed that there was a link between tobacco use and lung cancer: people diagnosed with lung cancer were more likely to have smoked than people without lung cancer (Müller, 1940; Proctor, 2012). Later in 1950, multiple major epidemiological studies confirmed that cigarette smoking caused lung cancer (de Groot et al., 2018; Doll & Hill, 1950; Proctor, 2012; Ernest L. Wynder & Graham, 1950). In 1964, a report by the USA Public Health Service came to several crucial conclusions: (1) smoking elevates age-specific mortality by 70% in males and by a smaller extent in females; (2) the duration of smoking and the number of cigarettes per day (intensity) are positively associated with lung cancer risk; and (3) average male smokers have about 9 to 10-fold increased risk than male non-smokers, and the figure is 20 times higher in heavy smokers (U.S. Public Health Service & National Clearinghouse for Smoking Health, 1972).

Overall, ever smoking of any tobacco product (e.g. cigarettes, pipes and cigars) increased lung cancer risk by 5.50-fold (random-effects relative risk [RR] = 5.50; 95% CI, 5.07-5.96) (Lee, Forey, & Coombs, 2012). Compared to non-smokers, current-smokers were at 8.43 times higher risk (random-effects RR = 8.43; 95% CI, 7.63-9.31), and former-smokers had a relatively smaller risk than current-smokers (random-effects RR = 4.30; 95% CI, 3.93-4.71). Furthermore, the effect magnitude of current smoking was greater for squamous-cell carcinoma (random-effects RR = 16.91; 95% CI, 13.14-21.76) than adenocarcinoma (random-effects RR = 4.21; 95% CI, 3.32-5.34). However, smoking pipe/cigar only did not show a significant effect on the risk of adenocarcinoma (random-effects RR = 0.93; 95% CI, 0.62-1.40) (Lee, Forey, & Coombs, 2012).

Exposure to secondhand smoke (SHS) imposes a risk of lung cancer to non-smokers as well (American Cancer Society, 2015). In one study, exposure to SHS from male smokers posed a higher risk to non-smoking female partners with pooled relative risk (RR) of 1.27 (95% CI, 1.17-1.37) (Taylor, Najafi, & Dobson, 2007). In addition, a dose-response effect could be found in the association, which meant greater exposure to SHS is related to higher lung cancer risk (Taylor et al., 2007). And more recently, exposure to SHS at home during adulthood increased one's risk of lung cancer by 26% (pooled RR = 1.26; 95% CI, 1.09-1.46) (Hori, Tanaka, Wakai, Sasazuki, & Katanoda, 2016).

2.6.2.2 Lack of Physical Activity

Leisure-time physical activity (PA) is beneficial to health on many levels, one of them is reducing lung cancer risk. Recreational physical activity lessened the risk by 24% (pooled RR = 0.76; 95% CI, 0.69-0.85). Such protective effects were consistent across histological types, including adenocarcinomas (pooled RR = 0.80; 95% CI, 0.72-0.88), squamous-cell lung cancers (pooled RR = 0.80; 95% CI, 0.71-0.90) and small cell lung cancers (pooled RR = 0.79; 95% CI, 0.66-0.94). Effect of PA was statistically significant in current (pooled RR = 0.77; 95% CI, 0.72-0.83) and former-smokers (pooled RR = 0.77; 95% CI, 0.69-0.85), but significance did not appear in never-smokers (pooled RR = 0.96; 95% CI, 0.79-1.18) (Brenner,

Yannitsos, Farris, Johansson, & Friedenreich, 2016). PA worked favorably in risk reduction for ever-smokers, but the biological mechanism was still uncertain (Alberg et al., 2013). Besides, a lack of PA (i.e. an indicator of sedentary lifestyle) was associated with other unhealthy behaviors, like tobacco use. For this reason, potential residual confounding by cigarette smoking could not be ruled out (Alberg et al., 2013). Along with PA, body weight may also play a role. A large study in Norway evaluated the association between PA, smoking status, BMI, and lung cancer risk among female participants (Borch, Weiderpass, Braaten, Hansen, & Licaj, 2018). The study suggested a stronger association between higher PA level and lower risk in current and former-smokers, and in normal weight and overweight participants, but no significant association was found in the obese group (Borch et al., 2018).

2.6.2.3 Body Mass Index and Obesity

Obesity is a health problem characterized by excessive body fat accumulation. Body mass index (BMI), waist circumference, waist-to-hip ratio, and percentage of body fat are accessible methods to determine obesity status (Egom et al., 2018). Measurement of BMI together with other parameters provide more accurate information than BMI alone (Egom et al., 2018). However, because data used in this study was collected for another purpose, the current study relied on BMI as the sole indicator of obesity.

Body mass index (BMI) is a metric for delineating anthropometric height and weight characteristics in adults (Nuttall, 2015). It is calculated as weight in kilograms divided by height in meters squared. According to Health Canada, BMI is categorized into underweight (less than 18.5 kg/m^2), normal weight (18.5 - 24.9 kg/m^2), overweight (25.0 - 29.9 kg/m^2), obese class I (30.0 - 34.9 kg/m^2), obese class II (35.0 - 39.9 kg/m^2), or obese class III (40 kg/m^2 or greater) (Government of Canada, 2004). In 2008, Renehan et al. examined the effect size of the association between BMI and 20 cancer types by sex (Renehan et al., 2008). For lung cancer, every 5 kg/m^2 increase in BMI, there are diminutions in the risk by 24% (RR = 0.76; 95% CI, 0.70-0.83) in males, and 20% (RR = 0.80; 95% CI, 0.66-0.97) in females.

Therefore, a higher BMI is associated with a significantly lower incidence of lung cancer in both sexes (Renehan et al., 2008). More recent meta-analyses endorse the protective effect of excessive body adiposity on lung cancer risk (Duan et al., 2015; Yang et al., 2012).

The inverse association between higher BMI and lower lung cancer risk is found in many observational studies. For instance, Kollarova et al. (2008) found that overweight and obesity groups had a lower lung cancer risk than normal weight group after adjusting for age, sex, and smoking status. Duan et al. (2015) found that participants who were underweight had 1.31 times higher lung cancer risk compared to people had normal weight (RR = 1.31; 95% CI, 1.10-1.57), but a significant relationship was not observed in former (RR = 1.40; 95% CI, 0.82-2.36) or non-smokers (RR = 1.18; 95% CI, 0.90-1.54). The pooled RRs of 0.82 (95% CI, 0.77-0.86) and 0.78 (95% CI, 0.74-0.83) were found in overweight and obesity groups, respectively. The inverse relationship remained significant after stratification by smoking status in both levels (Duan et al., 2015). Yang et al. (2012) also identified an association between excessive body weight (BMI greater than 25 kg/m^2) and lower incidence of lung cancer, and they recognized that the strengths of the association were different by smoking status (Yang et al., 2012). A stronger protective effect against lung cancer was found in current-smokers with a RR of 0.63 (95% CI, 0.57-0.70), while effect magnitudes in former (RR = 0.73; 95% CI, 0.58-0.91) and non-smokers (RR = 0.83; 95% CI, 0.70-0.98) were smaller (Yang et al., 2012). Smoking might have confounded the relationship between BMI and the risk of lung cancer, because tobacco use has links with body weight (Sanikini et al., 2018; Yang et al., 2012). On the other hand, the effect of BMI remains significant after adjusting for smoking associated explanatory variables in the lung cancer screening model $PLCO_{m2012}$ (Tammemägi et al., 2013).

2.6.2.4 Socioeconomic Status (SES)

In one study (Torre, Siegel, & Jemal, 2016), 27.9% of people under the poverty threshold used tobacco products, and lung cancer is more frequently found in poor and less educated people in United

States (Alberg et al., 2013; de Groot et al., 2018). Even though smoking rates are negatively correlated with income, investigators also found a correlation between lung cancer incidence and lower socioeconomic status (SES) independent of tobacco use (Dalton et al., 2011; Sidorchuk et al., 2009). This denotes that low SES may also be associated with encountering harmful environmental factors, for instance, hazardous occupational exposure and housing conditions can impose additional adverse effects on people's health (de Groot et al., 2018). Education, another indicator of SES, also relates to the likelihood of lung cancer. Compared to people with a university degree, the relative risk of lung cancer was 1.33 (95% CI, 1.24-1.43) times higher for those with only a postsecondary diploma, 1.96 (95% CI, 1.85-2.08) for a secondary school diploma, and 2.52 (95% CI, 2.38-2.68) for less than secondary school diploma in both sexes (Mitra, Shaw, & Tjepkema, 2015).

2.6.2.5 History of Pulmonary Diseases

Inflammation caused by pulmonary diseases make people more liable to lung carcinogenesis. The positive association between level of a chronic inflammation marker CRP and lung cancer (LC) risk indicates a possible association exists between pulmonary inflammation and lung cancer formation (Chaturvedi et al., 2010; McCarthy et al., 2012). Inflammation in lung tissues relates primarily to respiratory conditions such as chronic obstructive pulmonary disease (COPD), pneumonia, and tuberculosis; their association with lung cancer are discussed here individually (Brenner, McLaughlin, & Hung, 2011; McCarthy et al., 2012).

Chronic obstructive pulmonary disease (COPD) refers to lung disease consisting of two main types, emphysema and chronic bronchitis (Durham & Adcock, 2015; Stark, 2013). The main determinant of COPD, cigarette smoking, is primarily responsible for lung cancers as well (Durham & Adcock, 2015; Stark, 2013). On the other hand, COPD represents a lung inflammatory pathway that may be different from an exclusive tobacco carcinogen pathway, so it may modify the risk of lung cancer. Moreover, the inflammation and remodeling caused by COPD are precursory indications of lung cancer, and airflow

constriction among smokers increased their risk of lung cancer by five times in comparison to that of smokers with normal respiratory function (Durham & Adcock, 2015; Young & Hopkins, 2010). Compared to people with normal lung function, the annual lung cancer incidence was doubled in people with COPD (Young et al., 2015). Patients who suffered from emphysema were more likely to die from lung cancer, even though they had never been active smokers (Brenner et al., 2011; Turner, Chen, Krewski, Calle, & Thun, 2007). A pooled analysis of seven case-control studies in Canada and Europe found that males with previous diagnoses of chronic bronchitis had 1.33 (95% CI, 1.20-1.48) times higher odds of developing lung cancer (Denholm et al., 2014). And a preceding emphysema increased the odds by 50% in males (95% CI, 1.21-1.87).

Pneumonia is an infection and inflammation of one or both lungs, and the infection can be caused by viruses, bacteria, fungi, and parasites (CDC, 2018; Cheepsattayakorn & Cheepsattayakorn, 2014). A meta-analysis of 22 studies investigated the association between an earlier diagnosis of pneumonia and lung cancer risk after adjusting for smoking status, and researchers found a higher risk of lung cancer without a sign of publication bias (overall RR = 1.43; 95% CI, 1.22-1.68) (Brenner et al., 2011). Among males, participants who had pneumonia 2 years or less before lung cancer diagnoses had higher odds of lung cancer than people free of the pulmonary condition after adjustment for age, smoking status and time since smoking cessation (OR = 3.31; 95% CI, 2.33-4.70) (Denholm et al., 2014). And another pooled study revealed that self-reported data on previous diagnosis of pneumonia increased lung cancer risk by 57% after stratifying by smoking status (RR = 1.57; 95% CI, 1.22-2.01) (Brenner et al., 2012).

Tuberculosis (TB) is a transmissible infectious bacterial disease. TB can attack any part of the body, but lungs are almost always affected. A meta-analysis of 30 studies evaluated the relationship between TB and lung cancer, and found 76% higher risk (RR = 1.76; 95% CI, 1.49-2.08) across all studies and 90% higher risk among never-smokers (RR = 1.90; 95% CI, 1.45-2.50) (Brenner et al., 2011). A pooled

analysis of 17 studies in Europe and North America also found that people previously diagnosed with TB were at a higher lung cancer risk independently of tobacco use, with a relative risk of 1.48 (95% CI, 1.17-1.87) (Brenner et al., 2011).

2.6.2.6 Diet and Nutrition

Thirty percent of all cancers can be explained by diet, and poor dietary intake can contribute to the development of lung cancer (Dela Cruz et al., 2011; Willett & Trichopoulos, 1996). In Canada, the estimated proportions of lung cancer incidence in the population that could be attributed to low fruits consumption and low non-starchy vegetables consumption were 11.1% and 5.3%, respectively (Poirier et al., 2019) A diet rich in fruits and vegetables is believed to protect high risk populations against lung cancer. People who consume the highest amount of fruits and vegetables had a 14% lower risk of lung cancer (RR = 0.86; 95% CI, 0.78-0.94) in comparison to the people with lowest intake; and such an inverse association was found marginally significant in current-smokers (RR = 0.90; 95% CI, 0.81-1.00), but not in never or former-smokers (Vieira et al., 2016). A significant dose–response association between each 100 grams increase in fruits and vegetables intake per day reduced the risk by 4% (RR = 0.96; 95% CI, 0.94-0.98). The risk can be reduced up to 27% by raising consumption to 400 grams per day, but no additional benefit was attained beyond this amount (Vieira et al., 2016).

Some studies group fruits and vegetables into separate classes and examined their influence on lung cancer risk, such as cruciferous vegetables which includes broccoli, cabbage, kale and so on (Alberg et al., 2013). Compared to individuals in the lowest category of total cruciferous vegetables consumption, the risk among people in the highest consumption category was 23% lower in case-control studies (random-effects pooled odds ratio = 0.77; 95% CI, 0.68-0.88) and 17% lower in cohort studies (pooled relative risk = 0.83; 95% CI, 0.62-1.08) (Lam et al., 2009). And the inverse association remains significant after stratification by smoking status; the ORs were 0.49 (95% CI, 0.27-0.92) in former and 0.52 (95% CI, 0.29-0.95) in current-smokers (Lam et al., 2010).

Humans obtain vitamin A from animal (retinol) and vegetable (carotenoid) sources, but only the vegetable component has potential protective effects against lung cancer (World Cancer Research Fund/American Institute for Cancer Research [WCRF/AICR], 2019). Since fruits and vegetables have shown a beneficial effect against lung cancer, researchers postulate that the dietary micronutrient carotenoid might have the potential to prevent lung cancer and reduce the risk. Many researchers attempt to understand the role of carotenoid in the etiology of lung cancer (Alberg et al., 2013). Gallicchio et al. conducted a systematic review investigating the effect of carotenoids on lung cancer development (Gallicchio et al., 2008). Dietary consumption and pre-diagnostic serum concentration of total carotenoids were associated with a 21% (pooled RR = 0.79; 95% CI, 0.71-0.87) and 30% (pooled RR = 0.70; 95% CI, 0.44-1.11) lower lung cancer risk in the highest exposure category in comparison with the lowest one after adjusting for smoking (Gallicchio et al., 2008). Results suggest an inverse association between carotenoids and lung cancer, although one of them is statistically insignificant. This insignificance may be a consequence of high collinearity between the carotenoids measurement and the intake of fruits and vegetables, or a residual confounding by cigarette smoking (Gallicchio et al., 2008).

Vitamins C and E (alpha-tocopherol) may also have protective effects (Dela Cruz, Tanoue, & Matthay, 2011). Yong et al. (1997) evaluated the relationship between dietary and supplemental intake of vitamin E, C, A, and lung cancer incidence in the First National Health and Nutrition Examination Survey epidemiologic follow-up study. The results showed that participants in the highest quartile of vitamin C intake had 34% lower risk of lung cancer compared to people in the lowest quartile (RR = 0.66; 95% CI, 0.45-0.96). In current-smokers with the lowest tertile of pack-years of smoking, the highest quartile of vitamin E intake lowered the risk of lung cancer by 64% in comparison with the risk in the lowest quartile (RR = 0.36; 95% CI, 0.16-0.83) (Yong et al., 1997).

On the contrary, studies of several large randomized control trials (RCTs) found that supplemental beta-carotene, a provitamin A carotenoids, increased lung cancer risk in smokers in the

Carotene and Retinol Efficacy Trial (CARET) (Goodman et al., 2004). Researchers found that high-dose beta carotene and retinol supplements increased the risk of lung cancer. The study recruited high risk people who had been extensively exposed to asbestos or had at least 20 pack-years of cigarette smoking history. In the intervention group, participants took a combination of 30 mg of beta-carotene and 25,000 IU of retinol daily. The study stopped ahead of schedule as the findings showed a harmful effect of the intervention. Compared to the placebo group, the risk of lung cancer in the intervention group was 12% higher (RR = 1.12; 95% CI, 0.97-1.31) (Goodman et al., 2004). The Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study examined the effect of daily supplemental alpha-tocopherol (50 mg), beta-carotene (20 mg), alpha-tocopherol plus beta-carotene, and placebo for 5 to 8 years on the incidence and mortality of cancers among male smokers aged 50 to 69 years old (Virtamo et al., 2003). Analyses revealed that the risk for lung cancer incidence was 6% higher in recipients of supplemental beta-carotene (RR = 1.06; 95% CI, 0.94-1.20) in comparison to recipients of the placebo (Virtamo et al., 2003).

The roles of soy foods and isoflavones in lung cancer risk has been studied as well. Yang and colleagues conducted a meta-analysis of eight case-control and three prospective cohort studies to evaluate the association between soy intake and lung cancer. Most of the individual studies controlled for smoking, secondhand smoke (SHS), energy intake, BMI, physical activity, fruits and vegetables intakes, alcohol drinking, and age. They found a 23% reduction in lung cancer risk with high intake of soy foods (overall RR = 0.77; 95% CI, 0.65-0.92) (Yang, Va, Wong, Zhang, & Xiang, 2011). Isoflavones, including genistein, daidzein, and glycitein, are found mainly in soy and soy food products. To evaluate the association between plasma Isoflavones and lung cancer risk, Shimazu et al. carried out a case-control study nested in a prospective cohort study. The odds ratio of lung cancer in the highest quintile compared to the lowest quintile of plasma genistein concentration was 0.31 (95 % CI, 0.12-0.86; *P* for

trend = 0.085) after adjusting for family history of lung cancer, pack-years of smoking among current-smokers, SHS, use of exogenous female hormones, and fruits and vegetables intake.

A report by World Cancer Research Fund and American Institute for Cancer Research in 2018 has summarized effects of a number of lifestyle factors on risk of lung cancer. The report concludes that there was a strong evidence for the positive effect of drinking water containing arsenic and taking high-dose beta-carotene supplements on lung cancer risk. The harmful effect of consuming red meat, processed meat, and alcoholic drinks were supported by some evidence (WCRF/AICR, 2019). The results showed that for every 100 grams of red meat consumption per day increase, there was an increased risk of lung cancer by 22% (RR = 1.22; 95% CI, 1.02-1.46). Haem iron in red meat is proposed to result in the production of free radicals. Red meat contains heterocyclic amines and polycyclic aromatic hydrocarbons when prepared at high temperatures. The dose-response meta-analysis revealed that every 50 grams increment of processed meat consumption per day increased the lung cancer risk by 14% (RR = 1.14; 95% CI, 1.05-1.24) (WCRF/AICR, 2019). N-nitroso compounds in processed meat and produced in the stomach are postulated to cause genetic mutation and cancer. Every 10 g increase in the daily consumption of ethanol was related to 3% higher lung cancer risk (RR = 1.03; 95% CI, 1.01-1.05). There are a number of suspected mechanisms of carcinogenesis: the carcinogenic metabolites of alcohol, increasing penetration of carcinogenic molecules into mucosal cells, and a lack of essential nutrients in diet. The report also summaries effects of several risk and beneficial factors according to the smoking status (WCRF/AICR, 2019). Every 40 mg increase in the vitamin C intake from food lowers lung cancer risk by 13% in current-smokers (RR = 0.87; 95% CI, 0.79-0.96). This could be possibly due to the effect of vitamin C on free radicals and reactive oxygen molecules, stimulating the immune system, and inhibiting production of carcinogens. In never-smokers, foods containing isoflavones may lower the risk of lung cancer. Risk was 12% lower in people with the highest consumption of food containing isoflavones in comparison to the risk in the lowest consumers (RR = 0.88; 95%, 0.79-0.99). After

stratification by smoking status, the highest consumers of isoflavones in never-smokers have 34% lower risk than the lower consumers (RR = 0.66; 95% CI, 0.51-0.84).

2.6.2.7 Other Modifiable Risk Factors

The effects caused by other smoking products, like marijuana and electronic nicotine delivery systems (also known as electronic cigarettes or e-cigarettes) may also relate to lung cancer. Other environmental risk factors, such as radiation, asbestos, and poor air quality, may exacerbate the development of lung cancer as well. And people infected with human immunodeficiency virus (HIV) may be at a higher risk than people free of HIV infection (Alberg et al., 2013; de Groot et al., 2018).

2.6.3 Explanatory Variables of Interest

In the current study, medical history includes two factors: BMI at age 50 and self-reported COPD. Age at study entry, sex, race/ethnicity, and education are main sociodemographic characteristics. Assessments of smoking exposures include smoking status (never, former, and current-smokers), intensity (number of cigarettes smoked per day), duration (number of years smoked), and time to quit in former-smokers (number of years since stopped smoking cigarettes). In the $PLCO_{m2012}$ prediction model, these factors demonstrated significant effects associated with lung cancer (Tammemägi et al., 2014, 2013). These variables are well established factors related to the outcome, therefore they were evaluated in the study.

2.7 Cigarette Smoking and Metabolic Markers

Cigarette smoking is associated with weight loss, so people gain weight over time during the cessation. The effect of continuing smoking on increasing level of leptin, a hormone that regulates energy expenditure, may explain lower body weight in smokers (Klok, Jakobsdottir, & Drent, 2007; Perkins & Fonte, 2002). For example, Targher et al. (2001) measured serum leptin concentrations in patients with type 1 diabetes mellitus and in controls who were matched on age, sex, BMI, blood pressure, and smoking status. They found that leptin levels were significantly lower in smokers in

comparison with non-smokers after stratification by diabetic status (Targher et al., 2001). Notwithstanding, Perkins and Fonte (2002) found no significant difference in means of leptin concentrations with respect to smoking status after controlling for BMI and age. More research is expected to evaluate the association between cigarette smoking and changes in leptin level in detail (Wang, Wang, & Wang, 2017).

In 2015, Cibickova et al. (2015) assessed the effect of smoking on lipid metabolism and insulin resistance. The measurement of C-peptide was obtained to describe one's insulin profile. The level of C-peptide with skewed distribution was analyzed with Mann-Whitney U-test (the Wilcoxon rank-sum test), and the results showed that the median of C-peptide level in non-smokers was significantly lower than the median in smokers (median: 2.11, IQR: 1.50-2.88 in non-smokers; median: 2.54, IQR: 1.76-3.53 in smokers; P-value < 0.001).

A raised hsCRP level is clinically related to higher risk of cardiovascular diseases (Kamath, Xavier, Sigamani, & Pais, 2015). Helmersson et al. (2015) investigated if hsCRP level differed by categories of smoking status: non-smokers, former, and current-smokers. The median in current-smokers was higher than former and non-smokers; however, the difference was not significant.

2.8 Obesity and Metabolic Markers

2.8.1 Obesity and Leptin

Leptin in the circulatory system can cross the blood-brain barrier, and elevated levels of leptin signal the brain to inhibit appetite and increase energy expenditure for maintaining a stable body weight (Klok et al., 2007). Nonetheless, leptin may fail to function normally. Many factors, such as prolonged exposure to high levels of circulating leptin and chronic inflammation, are partly responsible for leptin resistance, and leptin resistance is an important risk factor for obesity (Zhou & Rui, 2013). Many studies find a positive correlation between body fat and blood leptin level. For example, Kazmi et al. (2013) examined the association between obesity, BMI, and serum leptin concentrations in Rawalpindi. The

mean of leptin levels in the obese group (52.8 ± 24.6 ng/mL) was significantly higher than the mean in non-obese counterpart (6.3 ± 3.1 ng/mL) with a P-value of 0.001. And a moderate-to-strong correlation was detected between BMI and serum leptin concentration (r = 0.59; *P* = 0.001).

2.8.2 Obesity and Connecting Peptide (C-peptide)

Metabolic syndrome (MeS) is characterized by a combination of medical conditions, the central components include glucose intolerance, obesity, hypertension, and dyslipidaemia (Eckel, Grundy, & Zimmet, 2005). And MeS is a universally recognized risk factor for cardiovascular disease and type II diabetes mellitus (T2DM). Abdominal obesity or adiposity and insulin resistance are the most conspicuous clinical features of this syndrome (Deedwania & Gupta, 2006; O'Neill & O'Driscoll, 2015). Because of that, obesity status or BMI may potentially reflect people's metabolic health. Abdullah et al. (2012) grouped female university students into non-overweight (BMI < 25 kg/m^2) and overweight (BMI ≥ 25 kg/m^2) groups, and found that the mean of C-peptide in the overweight group (2.5 ± 0.7 ng/mL) was higher than that of the non-overweight group (2.1 ± 0.7 ng/mL) with a P-value of 0.02. Moreover, a weak positive linear relationship existed between C-peptide concentration and waist circumference among the overweight group with an r of 0.36 and a P-value of 0.02 (Abdullah, Hasan, Raigangar, & Bani-Issa, 2012). In the current study, as a result, a positive association was expected between BMI and C-peptide measurement.

2.8.3 Obesity and High-sensitivity C-reactive Protein (hsCRP)

An increase in CRP level could predict a higher risk of cardiovascular diseases in the future, and it is associated with a higher BMI as well (Visser M et al., 1999). A study comprised of 7,938 males and 8,678 females was conducted to verify if excess body fat was associated with elevated serum CRP concentration (i.e. values ≥ 0.22 mg/dL) (Visser M et al., 1999). Compared to males with normal weight, obese male participants had 2.13 (95% CI, 1.56-2.91) higher odds of having elevated CRP level after adjusting for age, race/ethnicity, smoking status, estrogen use (in females), inflammatory disease,

cardiovascular disease, and diabetes mellitus (Visser M et al., 1999). And females with obesity were 6.21 times (95% CI, 4.94-7.81) likelier to have elevated CRP concentration than their normal-weight peers (Visser M et al., 1999). The study by Clark et al. (2016) also found a positive association between BMI and CRP level with a P-value < 0.001, and the effect was stronger in females than in males.

2.9 Metabolic Markers of Interest

Obesity is accompanied by a series of metabolic changes which potentially act as a bridge between excessive body fat and risk of cancer. Some examples of the metabolic changes include altered concentrations of insulin, insulin-like growth factor-1 (IGF-1), sex-steroid hormones, and adipokines (Nimptsch & Pischon, 2016). From earlier studies of the PLCO Cancer Screening Trial data and biospecimens, certain metabolic markers showed significant associations with lung cancer (Goodwin et al., 2015). The expert team Dr. Azad, Dr. Goodwin, Dr. Stambolic, Dr. Dowling, Williams, Moore, Dr. Lohmann, and Dr. Shepherd at University of Toronto selected the most plausible metabolic markers from the candidate list with consideration of the pragmatic assessment of assay. Therefore, the current study assesses the roles of these metabolic markers: leptin, C-peptide, and hsCRP.

Leptin is the first discovered adipokine in 1994. It is produced in adipose tissue and mediated by leptin receptor. Leptin concentration in serum is correlated with the total fat mass and the volume of adipose cells. Leptin is important to control nutritional intake and to maintain energy balance. The level of leptin in males is lower than the level in females. Obesity, insulin stimulation, acute infections, and the effect of cytokines can result in upregulation of leptin production. A decrease in leptin can be a consequence of fasting, exposure to cold, testosterone, and beta-adrenergic agonists (Ntikoudi, Kiagia, Boura, & Syrigos, 2014).

C-peptide, also known as connecting peptide, is synthesized along with insulin in equimolar amounts. C-peptide is a marker of endogenous insulin secretion. Higher concentrations of C-peptide

may indicate an increased atherogenic activity and a higher risk for cardiovascular disease. Elevated C-peptide may also suggest an increased risk of cancers (Aleksandrova, Mozaffarian, & Pischon, 2018).

Positive and negative acute-phase reactants are proteins whose plasma concentration can increase or decrease by at least 25% during an inflammation (Epstein, Gabay, & Kushner, 1999). The acute-phase C-reactive protein (CRP) is a sensitive metabolic marker for systemic inflammation (Visser M et al., 1999). It is released into the blood by the liver following acute infections, inflammatory conditions, and trauma. Highly sensitive assay is a technologically optimized design that can detect low levels of C-reactive protein, and its measure is denoted as high-sensitivity C-reactive protein (hsCRP). CRP recognizes and removes many pathogens and abnormal cells by the humoral immune response. The concentration of CRP is substantially low in healthy people, but an inflammation in response to infections, autoimmune reactions, cardiovascular diseases, sepsis, and cancer can trigger a rapid increase in the level of CRP. The extent of the elevation is correlated with the severity of tissue injury or inflammatory reaction (Agassandian, Shurin, Ma, & Shurin, 2014).

2.10 Dietary Factors and Metabolic Markers

Food frequency questionnaires (FFQs) are commonly used in dietary studies to gather information on how often each food item is consumed. Tracking patterns of dietary intake can provide empirical evidence to analyze data related to risk of lung cancer. Hu et al. (1999) performed factor analysis to identify dietary patterns based on nutrient profiles and culinary use in two categories: Western or Prudent. The typical constitution of a Western pattern includes high consumptions of refined grains, processed meat, red meat, butter, high-fat dairy products, and eggs. While the Prudent dietary pattern consists of a large consumption of fruits, vegetables, legumes, whole grains, poultry, fish, and other seafood independent of BMI (Hu et al., 1999).

2.10.1 Dietary Factors and Leptin

The level of circulating leptin showed fluctuations at 24-hour intervals and also change regarding dietary circumstances: feeding increases leptin concentrations and a drop can be detected during fasting (Münzberg & Morrison, 2015). Ko et al. (2016) evaluated the relationship between diet quality, patterns, and metabolic markers including leptin and C-reactive protein (CRP) after adjusting for BMI. The results showed a greater positive relationship between Western diet score and total energy intake ($r = 0.85$, $P < 0.001$). On one hand, Western diet scores were positively associated with leptin concentration (standardized $\beta = 0.50$, $P = 0.008$). On the other hand, a negative association was found between Prudent diet scores and leptin level (standardized $\beta = -0.24$, $P = 0.009$). Moreover, there was a descending trend in leptin levels when the tertile of vegetables, fruits, legumes, eggs, and fish consumption went up (Ko et al., 2016).

2.10.2 Dietary Factors and C-peptide

Fung et al. (2001) examined the association of Western and Prudent dietary patterns and metabolic markers in relation to cardiovascular disease and the risk of obesity (Fung et al., 2001). The study found that people with higher Prudent scores had a lower insulin concentration, and a higher C-peptide level was detected in participants with greater Western diet scores. They also observed a significant positive correlation between the Western dietary pattern and log-transformed C-peptide concentration ($r = 0.31$, $P < 0.01$) after adjusting for age, smoking status, total energy intake, alcohol intake, physical activity, hours of television watching, and BMI (Fung et al., 2001).

2.10.3 Dietary Factors and hsCRP

The influence of diet on inflammatory markers have been documented by Smidowicz and Regula (2015). In another study by Fung et al., the association between dietary patterns and CRP level was evaluated (Fung et al., 2001). The means and standard errors of CRP in the first, third, and fifth quintile of Western diet were 1.7 ± 0.3, 1.9 ± 0.3, and 2.5 ± 0.3 mg/L (P for trend = 0.004). And a significant

positive correlation was observed between log-transformed CRP concentrations and Western dietary pattern scores (r = 0.22, *P* < 0.0001) independent of BMI (Fung et al., 2001).

Nettleton et al. (2006) examined the relationships between dietary pattern, inflammatory markers, and endothelial activity. The fats and processed meats dietary pattern involving oils, fats, processed meats, fried potatoes, desserts, and salty snacks were positively associated with CRP concentration (*P* for trend < 0.001). In contrast, an inverse association was identified between whole grains and fruits dietary pattern (whole grains, fruits, green leafy vegetables, and nuts) and plasma CRP level (*P* for trend < 0.001). Additionally, a diet pattern typified by high consumption of vegetables and vegetable oil, resulting in high intakes of antioxidant micronutrients and essential fatty acids, was significantly associated with 12% lower odds of having CRP level greater than 3 mg/L after adjusting for baseline BMI (OR = 0.88; 95 % CI, 0.78-0.98) (Julia et al., 2013).

2.11 Mechanism of the Relationship between Metabolic Markers, Metabolic Changes, and Lung Cancer

Leptin's angiogenic, mitogenic, and immunomodulatory effects have been stressed by many studies, which suggest that it has a potential to mediate and amplify the progression of carcinogenesis, such as tumor invasion and migration (Ntikoudi, Kiagia, Boura, & Syrigos, 2014; Ribeiro, Araújo, Lopes, & Medeiros, 2007). Nonetheless, the relationship between serum leptin level and lung cancer is indeterminate. Investigators hypothesize that serum leptin level is higher if patients have advanced lung cancer. As the disease progresses, less leptin is produced because cachexia induced body fat loss is more commonly found at advanced stages (Ntikoudi et al., 2014). A cohort study from the Third National Health and Nutrition Examination Survey (NHANES III) evaluated initial serum C-peptide concentration and lung cancer death (Hsu, Chang, Lin, Lin, & Caffrey, 2013). After adjusting for age and smoking status, higher serum C-peptide concentration was associated with increased mortality among females (HR = 2.65 per pmol/mL; 95% CI, 1.31-5.36). In a prospective cohort study, Allin, Bojesen, and Nordestgaard

(2009) measured baseline CRP level of 10,408 participants followed for up to 16 years. Participants with CRP levels greater than 3 mg/L at baseline were at a 2.2 higher risk of developing lung cancer in comparison with those who had CRP lower than 1 mg/L after adjusting for age, sex, smoking, alcohol consumption, BMI, and for women, also oral contraceptive therapy, menopausal status, and hormone replacement therapy (hazard ratio = 2.2; 95% CI, 1.0-4.6). A previous matched case-control study nested in the PLCO Cancer Screening Trial also identified that a higher circulating CRP level was associated with higher prospective lung cancer risk (Shiels et al., 2013).

2.12 Source Data — The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial

In 1993, the National Cancer Institute launched the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial to investigate disease-specific mortality reduction from screening tests for these four cancers compared with usual medical care practices in males and females. Secondary analyses included screening test operating characteristics, incidence, stage, survival of cancer cases, costs, risks, etiology, and natural history of disease (Jones & Hattersley, 2013). The study design is briefly summarized here, while an explicit description can be found in previous papers (Gohagan et al., 2000; Prorok et al., 2000).

2.12.1 Current Study Sample Data

A matched case-control study design nested in the intervention group of the prospective PLCO Cancer Screening Trial was used to evaluate associations between diet, BMI, metabolic markers, and lung cancer. Since secondary analysis was carried out using pre-existing data, study questions were restricted to available variables in the dataset. Among the intervention arm, eligible cases were participants who provided serum samples before their lung cancer diagnoses. Case and controls were matched on sex, age at entry (within five years of age), race/ethnicity, smoking status (never, former, and current), year of study entry, duration of follow-up, and study center. The observations were stratified by many aspects, so the five years of age variation leave more considerable cell sizes than a 2-

year matching age range. The control versus cases ratios were 3:1 in never-smokers and 2:1 in ever-smokers. The metabolic markers assessed from at least one year prior to diagnostic and non-fasting blood samples were used. Self-reported dietary information was collected at the study baseline and the fourth year of the trial by distributing food frequency questionnaires (FFQs). Only the latter dietary data were used for the current study because of superior quality. Answers for the FFQ, including an organized food list and a corresponding frequency section, reflects participants' usual frequency of each food item over the past 12 months. Because people tend to keep a same diet over time, the information on FFQs can represent individuals' long-term habits.

2.12.2 Study Centers & Ethical Considerations

The PLCO project recruited participants at ten screening centers across the United States (Gohagan et al., 2000; Prorok et al., 2000). In order to attain the total of 74,000 females and 74,000 males, each center required enrollment of about 5,000 to 30,000 individuals (Prorok et al., 2000). Local institutional review board approval was acquired to conduct the study at each screening center, and written informed consent were provided by all participants (Oken et al., 2011). For the current study, the appropriate Data Transfer Agreement has been signed off on by Brock University and the U.S. National Institutes for Health, and Brock University Research Ethics Board clearance has been received for secondary data analysis.

2.12.3 Recruitment and Participant Cohort

The PLCO Cancer Screening Trial was a two-arm randomized controlled trial. It recruited people between 55 to 74 years of age. People diagnosed with any of four cancers under study or were receiving treatments for any cancer other than basal cell or squamous cell skin cancer were excluded from the study. People who had surgical removal of one lung, the entire colon, or the entire prostate were not included (Oken et al., 2005). Targeted individuals were from the general population dwelling in the catchment area of each screening center, and mass mailing was the main recruitment strategy (Oken et

al., 2011). In both intervention and control arms, 37,000 males were screened for lung, colorectal, and prostate cancer, and 37,000 females were screened for shared-site cancers and ovarian cancer (Gohagan et al., 2000). Recruiters tried to obtain an appropriate minority representation in the overall participant population (Oken et al., 2011; Prorok et al., 2000). A total of 154,942 participants enrolled for the trial, with 77,465 in the intervention and 77,477 in the control arm (Oken et al., 2005).

2.12.4 Randomization and Screening

Participants were stratified by screening center, sex, and age, and block randomized to either an intervention or a control arm. After that, they were followed for at least 13 years until diagnosis of cancer, death ascertainment or end of study to determine if screening reduced disease-specific mortality (Gohagan et al., 2000; Prorok et al., 2000). People received usual medical care practice in the control group. In the intervention group, posterior-anterior chest X-ray and sigmoidoscopy were used to screen for lung and colorectal cancers for both males and females. Digital rectal examination and blood test for prostate-specific antigen were employed to screen for prostate cancer in males, and females were assessed with transvaginal ultrasound and cancer antigen 125 blood test to examine for ovarian cancer (Gohagan et al., 2000; Prorok et al., 2000). For the lung cancer screened group, ever-smokers received chest X-ray at entry and annually for three consecutive years, while never-smokers were assessed biennially (Prorok et al., 2000).

2.12.5 Diagnostic and Therapeutic Follow-up

A nodule, mass, infiltrate, or other abnormalities on radiographs were considered to be suspicious for lung cancer, and the individuals were referred to a primary health care provider of their choice (Oken et al., 2005; Prorok et al., 2000). And the individuals with positive screening results were encouraged to receive diagnostic evaluation and treatment (Prorok et al., 2000). Moreover, in order to avoid providing biased therapy to people in the screening group, the PLCO trial decided not to dictate guidelines for diagnosis or treatment (Prorok et al., 2000). Results of screening tests and diagnoses were

identified and documented, and participants diagnosed with PLCO cancers were for the most part believed to have received diagnostic and therapeutic procedures according to currently accepted practice for his/her age, stage of cancer, and medical conditions.

2.12.6 Endpoints

The primary endpoint of the PLCO project was death from any of four cancers, and the secondary endpoint included the incidence, stage of cancer, and survival of cases (Prorok et al., 2000). Conducting active and passive follow-up and cause-of-death review processes maximized the completeness of endpoint information (Prorok et al., 2000). In each of the 10 screening centers, the death review committee determined the underlying cause of death by assessing all available relevant records, such as death certificate, pathology forms, autopsy reports, and pathology slides (Prorok et al., 2000). And the documents were reviewed by experts who had no affiliation with the trial and were blinded to the randomized group membership of the deceased (Prorok et al., 2000).

2.12.7 Data Collection

A self-administered baseline questionnaire was used at study entry to obtain information on personal sociodemographic characteristics (race/ethnicity, marital status, and education), family history of cancer, personal medical history, smoking history (status, age started and ceased, and pack-year history), and cancer screening history in last three years (Oken et al., 2011, 2005; Prorok et al., 2000). Diet-related data were collected in a separate questionnaire, including information on consumption of foods, cooking techniques, and use of dietary supplements (Prorok et al., 2000). An additional dietary questionnaire was included for the intervention group in their fourth year and for the control group at study entry (Hayes et al., 2000). Biological samples were collected for evaluation of cancer antigen 125, prostate-specific antigen, and other metabolic markers (Hayes et al., 2000). Blood was collected from participants in the intervention arm and aliquoted into serum, plasma, erythrocytes, and leukocytes annually at their screening visit (Hayes et al., 2000). Except for the whole blood, all other blood samples

were stored at -70°C at the PLCO biorepository in Frederick, Maryland (Hayes et al., 2000). For the current study, a 100-microliter aliquot of each serum sample temporally most distant from lung cancer diagnosis were shipped on dry ice to Mount Sinai Hospital in Toronto for analysis by Dr. Azar Azad.

2.13 Matched Case-Control Study Nested in the PLCO Trial

In the current study, a nested case-control study design was used. It embedded a matched case-control study within an established cohort. This study design reduced some shortcomings associated with classic case-control studies, and it also incorporated additional virtues of cohort studies (Sedgwick, 2014). Selection and recall biases can jeopardize traditional case-control studies, and inability to establish a temporal direction between exposure and outcome makes it difficult to determine a causal relationship (Patten, 2015). In a nested case-control study, a cohort is followed over time, so it can demonstrate the temporality of exposure and disease. And recall bias was not a major problem as information on a series of exposures was collected before the outcome occurs. Moreover, the cohort where both cases and controls came from provided a well-defined source population, which made the control selection simple to implement. Therefore, the risk of selection bias was minimized by the design as well. Besides, the high response rate and close follow-up of the cohort study made selection of cases and controls straightforward and enabled investigators to accurately identify incident cases (i.e. new cases) (Patten, 2015).

CHAPTER 3 METHODS

3.1 Overview

Study Questions:

1. Are recognized daily frequency of fruits and vegetables and supplemental beta-carotene intake associated with lung cancer? And do the effects differ by sex and histological type of lung cancer?

2. Is there an association between BMI and lung cancer? What are the relationships between dietary factors and BMI? Does BMI partly explain the effects of dietary risk factors associated with lung cancer? And do interactions exist?

3. Do metabolic markers affect the likelihood of lung cancer? What are the relationships between dietary factors and metabolic marker concentrations? Do relationships between BMI and metabolic markers exist? Do metabolic markers explain the effect of BMI on lung cancer likelihood? Do metabolic markers explain the effects of dietary factors on lung cancer likelihood? Are there interactions? How do sex and age interact with dietary factors, BMI, and metabolic markers in the observed associations?

To describe the associations between dietary factors, BMI, metabolic markers, and lung cancer, three study objectives have been proposed. For study objective one, dietary factors associated with lung cancer were selected by implementing prior knowledge and studying known lung cancer dietary risk explanatory variables in the dataset, and their relationships with lung cancer were evaluated using logistic regression. The associations in the pathway between Diet-BMI-lung cancer were evaluated for study objective two. Linear and logistic regressions were employed to describe relationships of Diet-BMI and BMI-lung cancer, respectively. Then the pathway between Diet-Metabolic markers-lung cancer was examined to address study objective three. Associations of Diet-Metabolic markers and BMI-Metabolic markers were described using linear regression, and logistic regression evaluated the relationship between metabolic markers and lung cancer. Analytical procedures were performed in the statistical

software Stata/IC 14.2 (StataCorp, College Station, Texas). Two-tailed significance levels less than 0.05 were considered significant for hypotheses testing.

3.2 Data Preparation

Every explanatory variable under investigation was assessed for missing data and outliers. Usually, a small portion of missing (5% or less) is inconsequential. If the percentage of participants with complete information on all variables is lower than 90%, then the results may be biased. In this case, the mechanisms and patterns of missing data should be investigated (Dong & Peng, 2013). Based on an underlying mechanism, an incompleteness can be classified into one of three types: missing completely at random, missing at random, and missing not at random (Little, Jorgensen, Lang, & Moore, 2014). Disregarding the observations which have missing values is a way to handle the issue, but a loss of information from these observations and lower study power are major drawbacks. A better approach to manage missing values is data imputation that can make an attempt to obtain unbiased parameter estimates (Graham, Olchowski, & Gilreath, 2007). Multiple imputations (MI) offers a set of plausible values for each missing value of covariates, and it is suitable for many types of missing information. Recent studies recommended that the number of imputation should be similar to the percentage of incomplete observations to gain a good reproducibility level of the results (Pedersen et al., 2017). For instance, 20 imputed datasets are required when there is 20% of data missing. And all explanatory variables involved in the types of missing and outcome variable are included in the imputation model.

Carefully tackling outliers is also crucial, as they are extreme values distinctly different from the rest of the data. Outliers can be visually detected using the Stata command *scatter* and command *graph box* to generate scatter plots and boxplots. And another Stata command *extremes* gives values of Cook's distance, any observation with Cook's distance greater than 4/N can potentially influence the results.

3.3 Candidate Variables

Candidate explanatory variables were selected based on formerly identified risk factors for lung cancer, current knowledge, and previous studies of the PLCO Cancer Screening Trial. Candidate dietary factors and metabolic variables are summarized in Table 4. Other explanatory variables and potential confounders are listed in Table 5 with corresponding descriptions. For study objective one, two, and three, lung cancer status (yes/no) is the response variable. The sociodemographic characteristics include age at study entry, race/ethnicity, and education (7 categories) (Tammemagi et al., 2011). A positive COPD stands for the presence of emphysema or/and chronic bronchitis. Positive self-reported COPD may relate to higher odds of getting lung cancer. Therefore, education and history of COPD are potential confounders in the study.

Among lung cancer patients, the outcome variable can be split into two complementary groups to explore effects on different lung cancer types: non-small cell lung cancer (NSCLC) versus small cell lung cancer (SCLC). Another comparison can be drawn among adenocarcinomas (ADs), squamous-cell carcinomas, and Other NSCLC. The rationale behind the categorization was explained in section 2.2.2.

Table 4. Candidate Dietary Variables and Metabolic Markers associated with Lung Cancer

Variable Categories	Variable Names	Additional Descriptions
Dietary Factors (n=7)	Food energy from diet	In kcal/day
	Sum of daily frequency of fruits and vegetables	Sum of natural log transformed "daily frequency of fruits, including beverages" and "daily frequency of vegetables, including juice excluding fried potatoes, adjusted for ketchup, onion, and garlic amounts"
	Red meat consumption	Hamburgers, steak, pork chops, bacon, and regular sausage, in g/day
	Vitamin C from diet	In mg/day
	Total beta-carotene	Beta-carotene from diet and current supplements only, in mcg/day
	Supplemental beta-carotene	Supplemental beta-carotene from current single and multi-vitamins only, in mcg/day
	Total isoflavone	In mg/day
Metabolic Markers (n=3)	Log transformed Leptin	In ln(ng/mL)
	Log transformed C-peptide	In ln(pmol/L)
	Log transformed hsCRP	In ln(mg/L)

Table 5. Candidate Variables and Potential Confounders for Evaluating Associations with Lung Cancer

Variable Categories	Variable Names	Additional Descriptions
Sociodemographic (n=2)	Race/ethnicity	White, Non-Hispanic Black, Non-Hispanic Hispanic Asian Pacific Islander American Indian/Alaskan Native
	Education	Less than 8 years 8 to 11 years 12 years or completed high school (referent group) Post high school training other than college Some college College graduate Postgraduate degree
Medical History (n=2)	BMI at age 50	In kg/m^2
	Self-reported COPD	Presence of emphysema or/and chronic bronchitis
Smoking Exposure (n=3)	Number of cigarettes smoked per day	Numeric
	Duration smoked cigarettes	In years
	Number of years since stopped smoking cigarettes (in former smokers)	In Years
Confirmed Lung Cancer	Lung cancer type (small or non-small cell lung cancer)	Small cell cancer Non-small cell lung cancer Adenocarcinoma Squamous cell carcinoma Other Non-small cell lung cancer

3.4 Descriptive Statistics

Frequency, percentage, and proportion of each category were assessed in categorical variables, such as sex, smoking status, lung cancer status, and histology. The distributions, measures of central tendency, and variability of quantitative variables, including dietary factors, BMI, and metabolic markers concentrations were evaluated. Means and standard deviations were calculated for normally distributed quantitative variables, and Student's independent sample t-tests were employed to compare the differences in means of these variables between two groups. Whereas, medians and interquartile ranges

were computed for quantitative variables with skewed distribution. The Wilcoxon rank-sum test (also known as the Mann–Whitney test) is a non-parametric equivalent of the independent t-test, and it can identify if the two populations where the two samples are selected from have the same distribution (Field, 2009). After summarizing study covariates overall, these analyses were carried out stratifying by lung cancer. Chi-square tests of independence evaluated the differences in distribution between two categorical variables; when any expected frequency was less than five, then Fisher's exact test were used. And the direction and strength of relationships between metabolic markers concentrations were assessed using correlation.

3.5 Univariate Analysis

For study objective one, two, and three, univariate conditional logistic regressions were carried out separately to evaluate the associations between dietary factors, BMI, metabolic markers (leptin, C-peptide, and hsCRP), and lung cancer odds. Correlation and linear regression methods evaluated associations between dietary factors, BMI, and metabolic markers.

3.6 Multivariable Analysis

Multivariable conditional logistic regressions included the explanatory variables and possible interaction terms in ever-smokers. The non-candidate inclusion criterion was set to 0.15, explanatory variables in univariate analyses show a P-value less than 0.15 were evaluated in later multivariable analyses (Bursac, Gauss, Williams, & Hosmer, 2008).

3.7 Model Building Approach

3.7.1 Handling Quantitative Variables

The distributions of metabolic markers were highly skewed, so converting them to normal distributions was fundamental before including them as quantitative variables. The Stata command *gladder* generates multiple different transformations and depict them with graphics; then a proper transformation could be chosen from them by judging the distributions. The Stata command *qladder*

illustrated quantile-normal plots for each transformation, and the plots were used to evaluate the normal distribution of the transformed variables. Usually, a natural logarithm is a commonly used transformation for metabolic marker values (Grund & Sabin, 2010). Multivariable fractional polynomials (MFP) were used to handle non-linear associations between quantitative explanatory variables and lung cancer (Sauerbrei, Meier-Hirmer, Benner, & Royston, 2006).

3.7.2 Evaluating Collinearity

Collinearity or multicollinearity occurs when two explanatory variables are highly correlated with each other. Consequently, it leads to volatile model estimates of the coefficients and inflated standard error. The Stata command *vif* and *collin* computed variance inflation factor (VIF) and a tolerance value (1/VIF), and they represented how severe the variance is inflated. Furthermore, carefully considering the causal framework under investigation is more important to understand what causes the collinearity (Schisterman, Perkins, Mumford, Ahrens, & Mitchell, 2017). The correlations among the 3 metabolic markers were evaluated.

3.7.3 Assumption Checking

1) Independence of Errors: the observations of the data should not be related.

2) Linearity: a linear relationship exists between any quantitative independent variable and the logit of the dependent variable. Non-linear relationships were evaluated by adding a Stata prefix command *mfp* to run analysis with multiple fractional polynomial regressors.

3) Lack of Influential Outliers: too many outliers may compromise the overall accuracy of the model. Examining residuals (the difference between predicted and actual outcomes), diagnostic statistics and graphs could locate the outliers. Comparing the overall model fit and estimated beta coefficients with and without the outliers demonstrated how influential the outliers are, and one can decide to retain or remove them (Stoltzfus, 2011).

3.7.4 Model Selection

Prior knowledge and data should be considered to select model and variables (Greenland, 1989). Besides evaluating the relationships between dietary variables and lung cancer status, the existence of confounders, effect modifiers, and mediators need to be carefully examined. Both confounders and effect modifiers are extraneous variables. A confounder is a risk factor for lung cancer and also relates to diet, but it does not occur in the proposed causal pathway. Confounding variables distort the associations under investigation, therefore they should be carefully assessed and controlled (Patten, 2015). On the other hand, effect modification refers to a situation when the effect of an exposure on the outcome differs across the level of another exposure (Knol, Egger, Scott, Geerlings, & Vandenbroucke, 2009). In analytical phase of the study, stratification can unveil the effect of a categorical extraneous variable. To identify if a given exposure is a confounder is to quantify the effect of confounding. The first step is to compute the effect estimates both before and after adjusting for the potential confounder. Next, calculating the difference between the unadjusted and adjusted measures of association in percentage reveals the magnitude of confounding. When the difference is more than 15%, there is evidence of confounding. In this case, the confounder is important in the association and adjusted effect estimates should be reported. If the stratum-specific estimates are different from each other and the crude estimate falls in between, this implies an effect modifier. Both confounding and effect modification by quantitative variables can be evaluated using a multivariable regression modelling. An important confounder is kept in the regression model. To assess an effect modification, the suspected quantitative variable is included in the model as a cross-product term (interaction term) along with main effects. If the cross-product term between exposure and the covariate is significant, the covariate is an effect modifier. Otherwise, further assessment for confounding takes place by comparing adjusted and unadjusted estimates (Patten, 2015). A mediator displays an intermediate effect between diet and lung cancer, and it partially describes the association (Corraini, Olsen, Pedersen, Dekkers, &

Vandenbroucke, 2017). The portion between crude and adjusted estimates represents the effect of the mediator. As illustrated in Figure 1, BMI, leptin, C-peptide, and high-sensitivity C-reactive protein (hsCRP) are potential mediators.

Controls were matched to cases from PLCO data, and an unconditional analysis would not completely eliminate confounding by matching factors. In order to acquire unbiased results, conditional analysis should be carried out (Pearce, 2016). As mentioned before, explanatory variables with a P-value less than 0.15 in univariate analyses were included in multivariable models. Backward elimination was used for model selection: all explanatory variables were placed in the model initially, the contribution of each explanatory variable was calculated, and the significance values were compared to a removal or an inclusion criterion. The least influential explanatory variable or the one which meets the removal criterion was dropped. This process may be required multiple times to obtain a final model. Using backwards elimination minimized the issue of suppressor effects, which happens when an explanatory variable possesses a significant effect but only when another variable is kept consistent (Field, 2009).

3.8 Model Evaluation/Diagnostics

3.8.1 Numerical Measures

3.8.1.1 Measures of Fit

Using the Stata command *fitstat* gives the measures of fit for the purpose of model comparison. The model with smaller Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) is considered the better one.

3.8.1.2 Model Adequacy/Specification Error

The Stata command *linktest* can detect a specification error. The principle of *linktest* is that if no additional significant explanatory variables can be determined except by chance, then the model is properly specified. It rebuilds the model using the predicted value (_hat) and the predicted value squared (_hatsq) as the explanatory variables. The variable _hat should be statistically significant,

otherwise an insignificance denotes that the entire model is incorrectly specified. On the contrary, the variable _hatsq should be an insignificant explanatory variable. A significant _hatsq indicates one of following possibilities: important variables are not included in the first place, relevant variables are falsely removed from the model or modeled incorrectly, and the dependent variable is falsely specified. Other causes may be the lack of availability for inclusion, inappropriately transformed quantitative variables, or missed interactions.

3.8.2 Graphical Inspections

3.8.2.1 Plot of Standardized Residuals versus Observation Number

Pearson standardized residual can be used to visually inspect influential observations. Pearson chi-square residuals are the differences between observed and fitted values. When plotting standardized residual (y-axis) versus observation number (x-axis), values which are outside the range of negative and positive 1.96 on the y-axis are potentially influential observations.

3.8.2.2 Plot of Influential Observations/Analysis of Residuals Using Cook's Distances

Cook's distance of a variable depicts the change in all fitted values when the variable is removed; hence, a large value of Cook's distance indicates an influential observation. When plotting Cook's distance (y-axis) versus observation number (x-axis), influential values manifest themselves as peaks.

CHAPTER 4 RESULTS

Initially, the study sample included both never and ever-smokers. The descriptive characteristics of the pooled sample are included in the Supplemental Material Table S1. As the two groups were different in terms of biological mechanism and carcinogenesis, analyses were conducted separately. In never-smokers, none of the study exposures are significantly different between controls and cases (Table S2). For this reason, never-smokers are excluded from the subsequent analysis.

Data Preparation

The overall completeness of the dataset for variables of study interest is 88%. There is no missing variable in the outcome of interest. Twenty-four extreme values detected in explanatory variables were truncated to the maximum of the remaining observations (n=3) or set as missing values (n=21). All dietary factors and metabolic markers were natural log transformed (LN) according to the results in Stata's *ladder of power* except supplemental beta-carotene intake. The ordinal variable supplemental beta-carotene intake was rescaled to 1000 mcg/day from 1 mcg/day. To prevent multicollinearity, quantitative variables were centered on means or values close to means. Supplemental beta-carotene intake was centered on 500 mcg/day. LN daily frequency of fruits and vegetables was centered on 2 units. LN leptin concentration was centered on 1.8 ln(ng/mL). LN C-peptide level was centered on 7 ln(pmol/L). LN high-sensitivity C-reactive protein (hsCRP) concentration was centered on 2 ln(mg/L). BMI at age 50 was centered on 27 kg/m^2. Age at study entry was centered on 64 years. Smoking duration was centered on 30 years. Educational achievement was centered on level four (post high school training other than college).

Descriptive statistics and results of univariate logistic associations for ever-smokers are presented in Table 6. Table 7 provides a summary of lung cancer histological distribution in the study. Important correlations between explanatory variables are summarized in Table 8. The final multivariable conditional logistic regression model is presented in Table 9 followed by an assessment of collinearity

(Table 10). Assumptions were examined together with visual inspections (Figure 4 and 5). Measures of fit and model adequacy were diagnosed for the final model. Then the explanatory variables in the final model were evaluated in different histological subtypes of lung cancer.

4.1 Population Characteristics

The characteristics of ever-smokers are presented in Table 6. In ever-smokers, there were 1,685 controls and 855 lung cancer cases. Participants' ages at study entry ranged from 55 to 74 years (mean = 64.0; SD = 5.0) with 37.5% identified as female. The majority of participants (89.8%) were Non-Hispanic White. The similarities in age, sex, and race/ethnicity distribution between controls and cases reflect the matched case-control study design. About 67.9% of controls and 62.1% of cases had greater than high school education. There was an association between educational achievement and lung cancer ($P_c <$ 0.001). Specifically, higher educational level was related to lower odds of lung cancer ($P_{overall} < 0.001$). The mean BMI at age 50 in controls was 0.45 kg/m^2 higher than the value in cases, which indicates an inverse association with lung cancer ($P = 0.002$).

Among seven candidate dietary factors, significant mean differences were found between control and case groups for the LN daily frequency of fruits and vegetables, LN red meat consumption, LN supplemental beta-carotene intake, and LN vitamin C from diet. Out of 2,540 ever-smokers, 1375 (54.1%) were former-smokers. The similar proportion of current-smokers in controls and cases is expected as smoking status is one of the matching criteria in the study. Although a difference in smoking status does not exist, dissimilarities are identified in some detailed smoking characteristics. On average, cases smoke 5.7 cigarettes/day more and 6.9 years longer than controls. The average number of years since smoking cessation in controls was 10.1 years longer than in cases. The mean level of LN C-peptide in cases was 0.06 ln(pmol/L) higher than in controls. The average LN hsCRP concentration was 0.26 ln(mg/L) lower in controls than in cases. No significant difference was found in leptin concentration.

Univariate conditional logistic regressions were carried out for all explanatory variables and covariates, excluding sex, race/ethnicity, and smoking status. Univariate odds ratios, 95% confidence interval estimates, and P-values for unadjusted effects are shown in Table 6. The results of univariate analyses served as a guideline for multivariable analysis. Higher LN fruits and vegetables daily frequency, supplemental beta-carotene intake, LN vitamin C from diet, and LN isoflavone were associated with lower odds of lung cancer. Higher LN red meat consumption was related to higher odds of lung cancer. Higher BMI at age 50 demonstrated a protective effect against lung cancer. Higher LN pre-diagnostic C-peptide and hsCRP were associated with a higher likelihood of lung cancer. A trend towards significance suggests a potential inverse association between LN leptin level and lung cancer.

Furthermore, the distribution of lung cancer type is summarized in Table 7. Among 855 participants with lung cancer, 119 (13.9%) of them were diagnosed with small cell lung cancer (SCLC). In 736 people with non-small cell lung cancer (NSCLC), adenocarcinoma, including bronchiolo-alveolar carcinoma (BAC), contributed to 319 (37.3%) of them. Apart from 188 (22.0%) participants with squamous cell carcinoma, there were 229 (26.8%) who were diagnosed with Other NSCLC (in this study, including large cell carcinoma, other NSCLC, not otherwise specified carcinoma, and other/missing).

Table 6. Characteristic of Ever-smokers by Lung Cancer Status and Univariate Logistic Associations with Lung Cancer

Explanatory Variables	Controls (n = 1685)	Cases (n = 855)	*P*	Univariate Odds Ratio (95% CI; *P*)
Sociodemographic				
Age, year, mean (SD)	63.94 (4.99)	63.98 (4.97)	$P_t = 0.868$	1.169 (0.998–1.370; 0.053)
Sex, number (%)			$P_c = 0.962$	Controls matched to cases
Female	631 (37.45)	321 (37.54)		
Male	1054 (62.55)	534 (62.46)		
Race/ethnicity, number (%)			$P_e = 0.888$	Controls matched to cases
Non-Hispanic White	1514 (89.85)	768 (89.82)		
Non-Hispanic Black	109 (6.47)	56 (6.55)		
Hispanic	21 (1.25)	10 (1.17)		
Asian	29 (1.72)	15 (1.75)		
Pacific Islander	6 (0.36)	5 (0.58)		
American Indian/Alaskan Native	6 (0.36)	1 (0.12)		
Education, number (%)			$P_c < 0.001$	
Less than 8 years	18 (1.07)	11 (1.29)		1.032 (0.470–2.267; 0.937)
8 to 11 years	138 (8.21)	91 (10.64)		1.121 (0.812–1.548; 0.486)
12 years or completed high school	381 (22.67)	222 (25.96)		Referent group
Post high school training other than college	234 (13.92)	122 (14.27)		0.887 (0.670–1.174; 0.403)
Some college	385 (22.90)	220 (25.73)		0.997 (0.785–1.267; 0.980)
College graduate	258 (15.35)	113 (13.22)		0.769 (0.581–1.017; 0.066)
Postgraduate degree	267 (15.88)	76 (8.89)		0.492 (0.359–0.673; <0.001) <0.001(overall)
Medical history				
BMI at age 50, kg/m^2, mean (SD)	25.43 (3.52)	24.98 (3.43)	$P_t = 0.002$	0.964 (0.939–0.989; 0.006)
Self-reported COPD, number (%)			$P_c < 0.001$	
No	1531 (90.86)	713 (83.39)		
Yes	154 (9.14)	142 (16.61)		1.934 $_{\text{yes vs no}}$ (1.506–2.483; <0.001)

Abbreviations: BMI, body mass index (weight kg/height meter^2); CI, confidence interval; COPD, chronic obstructive pulmonary disease; CP, C-peptide; hsCRP, high-sensitivity C-reactive protein; IQR, interquartile range; LN, natural log transformed; *P*, P-value; P_c: P-value from Pearson's chi-squared; P_e, P-value from Fisher's Exact test; P_n, P-value from non-parametric test for trend across ordered groups (an extension of the Wilcoxon rank-sum test); P_t, P-value from t-test with unequal variances; SD, standard deviation.

(continued on the following page)

Explanatory Variables	Controls (n = 1685)	Cases (n = 855)	*P*	Univariate Odds Ratio (95% CI; *P*)
Dietary factors				
Food energy from diet, kcal/day, median (IQR)	2025.48 (1095.96)	1943.05 (1044.43)	$P_n = 0.106$	
LN Food energy from diet, mean (SD)	7.60 (0.42)	7.58 (0.43)	$P_t = 0.416$	0.908 (0.726–1.136; 0.397)
Fruits&Vegetables, daily frequency, median (IQR)	5.92 (4.05)	5.19 (3.82)	$P_n < 0.001$	
LN Fruits&Vegetables, mean (SD)	1.86 (1.17)	1.56 (1.22)	$P_t < 0.001$	0.787 (0.726–0.854; <0.001)
Red meat, g/day, median (IQR)	26.58 (36.41)	29.27 (38.96)	$P_n = 0.004$	
LN Red meat, mean (SD)	3.22 (0.97)	3.35 (0.95)	$P_t = 0.002$	1.158 (1.045–1.284; 0.005)
Supplemental beta-carotene, 1000 mcg/day, mean (SD)	0.51 (0.99)	0.42 (0.85)	$P_t = 0.021$	0.874 (0.789–0.968; 0.010)
Vitamin C from diet, mg/day, median (IQR)	149.95 (104.99)	129.78 (103.49)	$P_n < 0.001$	
LN Vitamin C from diet, mean (SD)	4.95 (0.60)	4.82 (0.60)	$P_t < 0.001$	0.708 (0.607–0.825; <0.001)
Isoflavone, mg/day, median (IQR)	0.40 (0.47)	0.36 (0.43)	$P_n = 0.019$	
LN Isoflavone, median (SD)	-0.98 (1.03)	-1.06 (0.96)	$P_t = 0.054$	0.899 (0.815–0.992; 0.033)
Smoking Exposures				
Smoking status, number of current-smokers (%)	772 (45.82)	393 (45.96)	$P_c = 0.943$	Controls matched to cases
Cigarettes per day, mean (SD)	23.63 (13.35)	29.29 (14.78)	$P_t < 0.001$	1.033 (1.026–1.040; <0.001)
Smoking duration, year, mean (SD)	32.31 (14.47)	39.16 (10.57)	$P_t < 0.001$	1.077 (1.065–1.088; <0.001)
Quit-time in former-smokers, year, mean (SD)	23.26 (12.51)	13.19 (10.45)	$P_t < 0.001$	0.929 (0.917–0.940; <0.001)
Metabolic markers				
CP (pmol/L), median (IQR)	752.14 (667.70)	824.85 (767.78)	$P_n = 0.011$	
LN CP, mean (SD)	6.65 (0.64)	6.71 (0.64)	$P_t = 0.034$	1.157 (1.012–1.323; 0.032)
hsCRP (mg/L), median (IQR)	8.50 (14.90)	11.65 (20.70)	$P_n < 0.001$	
LN hsCRP, mean (SD)	2.33 (1.01)	2.59 (1.04)	$P_t < 0.001$	1.323 (1.214–1.454; <0.001)
Leptin (ng/mL), median (IQR)	5.10 (6.90)	4.90 (7.30)	$P_n = 0.424$	
LN Leptin, mean (SD)	1.85 (0.77)	1.82 (0.79)	$P_t = 0.381$	0.939 (0.829–1.064; 0.325)

Abbreviations: BMI, body mass index (weight kg/height meter^2); CI, confidence interval; COPD, chronic obstructive pulmonary disease; CP, C-peptide; hsCRP, high-sensitivity C-reactive protein; IQR, interquartile range; LN, natural log transformed; *P*, P-value; P_c: P-value from Pearson's chi-squared; P_e, P-value from Fisher's Exact test; P_n, P-value from non-parametric test for trend across ordered groups (an extension of the Wilcoxon rank-sum test); P_t, P-value from t-test with unequal variances; SD, standard deviation.

Table 7. Distribution of Histological Subtypes of Lung Cancer in 2540 Ever-Smokers

	Lung cancer				Total
	Small cell lung cancer	Non-small cell lung cancer			
		Adenocarcinoma	Squamous cell carcinoma	Other NSCLC†	
Frequency (%)	119 (13.9)	319 (37.3)	188 (22.0)	229 (26.8)	855

† Other NSCLC in the current study includes large cell carcinoma, other NSCLC, not otherwise specified carcinoma, and other/missing.

4.2 Key Study Findings

Correlations were examined between quantitative explanatory variables and covariates. Explanatory variables that require greater attention are outlined in Table 8. LN food energy from diet is positively and moderately correlated with multiple explanatory variables: LN fruits and vegetables daily frequency, LN red meat consumption, LN vitamin C from diet, and LN isoflavone intake. The correlation coefficient between LN vitamin C from diet and LN fruits and vegetables daily frequency was as high as 0.8. LN isoflavone intake was potentially related to LN fruits and vegetables daily frequency and LN vitamin C from diet. As the BMI of participants at age 50 increased, their pre-diagnostic leptin concentration increased as well.

Table 8. Correlation Coefficient between Selected Explanatory Variables

Measures	1	2	3	4	5	6	7
1. BMI at age 50							
2. LN Food energy from diet							
3. LN Fruits&Vegetables		0.350					
4. LN Red meat		0.474					
5. LN Vitamin C from diet		0.489	0.846				
6. LN Isoflavone		0.307	0.353		0.305		
7. LN Leptin	0.291						

Study question 1. Are recognized daily frequency of fruits and vegetables and supplemental beta-carotene intake associated with lung cancer? And do the effects differ by sex and histological type of lung cancer?

The associations between dietary factors and lung cancer were evaluated using multivariable conditional logistic regressions. Sex was removed as no significance was found. After adjusting for age,

educational achievement, smoking intensity, and smoking duration, effect estimates of dietary factors are presented in Supplemental Material Table S3. Each one unit increase in LN daily frequency of fruits and vegetables was associated with 11.9% lower odds of lung cancer (OR = 0.881; 95% CI, 0.803-0.965; *P* = 0.007). Supplemental beta-carotene intake was associated with 13.3% lower odds of lung cancer (OR = 0.867; 95% CI, 0.773-0.971; *P* = 0.014). No interactions were found in the association between dietary factors and lung cancer.

Study question 2. Is there an association between BMI and lung cancer? What are the relationships between dietary factors and BMI? Does BMI partly explain the effects of dietary risk factors associated with lung cancer? And do interactions exist?

The univariate association between BMI and lung cancer was significant (Table 6, *P* = 0.006). A significant interaction effect was found between BMI and age in the association of BMI and lung cancer. The effect of BMI on lung cancer depended on the level of age (Table S4; OR = 1.007; 95% CI, 1.001-1.013; *P* = 0.024). Part of BMI effect might be explained by proposed dietary factors in the association. The strength of the inverse association between BMI and lung cancer decreased as age goes up.

The results of the multivariable linear regression with dietary factors predicting BMI are presented in Table S5. LN food energy from diet, LN fruits and vegetables daily frequency, LN red meat consumption, supplemental beta-carotene intake, LN vitamin C from diet, and LN isoflavone intake were significantly associated with BMI after adjusting for age, education, and smoking exposures. If other explanatory variables remain constant, each one ln(kcal/day) increase in LN food energy from diet corresponded to a 0.600 kg/m^2 increase in BMI (*P* = 0.011) on average. Each one unit (ln(frequency)) increase in fruits and vegetables daily frequency corresponded to a 0.407 kg/m^2 decrease in BMI (*P* = 0.002). Each one unit (ln(g/day)) increase in red meat consumption related to a 0.388 kg/m^2 increase in BMI (*P* < 0.001). Per one unit (ln(mg/day)) increase in vitamin C from diet was associated with a 0.875

kg/m^2 increase in BMI (P = 0.001). An assumption of the study was that participants did not change their diets over time. People with higher daily frequency of fruits and vegetables might consume more beverage and juice that are rich in both vitamin C and sugar. With a one unit (ln(mg/day)) increase in Isoflavone intake, on average BMI decreased by 0.243 kg/m^2 (P = 0.003). Although a number of dietary factors and covariates were considered in the model, effect modifications were not detected in the association.

After taking LN fruits and vegetables daily frequency and supplemental beta-carotene intake into consideration, higher BMI was significantly and inversely associated with lung cancer likelihood (Table S6). When covariates were kept constant, each one unit (kg/m^2) increase in BMI related to 3.9% lower odds of lung cancer on average (P = 0.017). Each one unit (ln(frequency)) increase in fruits and vegetables daily frequency corresponded to 12.1% lower odds of lung cancer (P = 0.006). Every 1000 mcg/day increase in supplemental beta-carotene intake was associated with 13.9% lower lung cancer odds (P = 0.010). The effects of fruits and vegetables daily frequency and supplemental beta-carotene intake on lung cancer did not differ by educational achievement, BMI at age 50, or smoking exposures.

Study question 3. Do metabolic markers affect the likelihood of lung cancer? What are the relationships between dietary factors and metabolic markers concentrations? Do relationships between BMI and metabolic markers exist? Do metabolic markers explain the effect of BMI on lung cancer likelihood? Do metabolic markers explain the effects of dietary factors on lung cancer likelihood? Are there interactions? How do sex and age interact with dietary factors, BMI, and metabolic markers in the observed associations?

The associations with dietary factors were examined for each of the three metabolic markers (Table S7a-c). P-values of adjusted effect estimates of fruits and vegetables daily frequency and LN red meat consumption on C-peptide, hsCRP, and leptin were all smaller than 0.10. Unlike the expected

negative relationship, LN fruits and vegetables daily frequency was positively associated with C-peptide (Table S7a). An inverse association was found between LN fruits and vegetables daily frequency and hsCRP, as well as leptin (Table S7b-c). The effect of LN fruits and vegetables daily frequency on C-peptide was different for male and female participants. For each one unit increase in fruits and vegetables daily frequency, a male participant had a 0.030 unit additional increase in C-peptide, but a female participant had a 0.027 additional decrease in C-peptide ($P_{interaction}$ = 0.016) (Table S7a). This indicates metabolic or hormonal physiological differences in males and females. LN red meat consumption was positively associated with all three metabolic markers. The effect of red meat consumption on hsCRP was different for sex and age (Table S7b). In addition, supplemental beta-carotene intake, LN vitamin C from diet, and LN isoflavone intake were significant explanatory variables of LN leptin level (Table S7c). The relationship between supplemental beta-carotene and leptin was non-linear. Unlike expected negative relationship, vitamin C from diet was positively associated with leptin concentration. Isoflavone intake was inversely associated with leptin.

The associations between BMI and C-peptide, and BMI and leptin were non-linear (Table S8a, c). The non-linearity detected between BMI at 50 and LN leptin concentration is illustrated in Figure 2. Increasing BMI was accompanied by a rapidly climbing in leptin level, then the trend decreases once BMI reaches 35 kg/m^2. A positive linear relationship existed between BMI and LN hsCRP, and a significant interaction of sex*BMI was detected in the association (Table S8b). The association between BMI and hsCRP differed by sex.

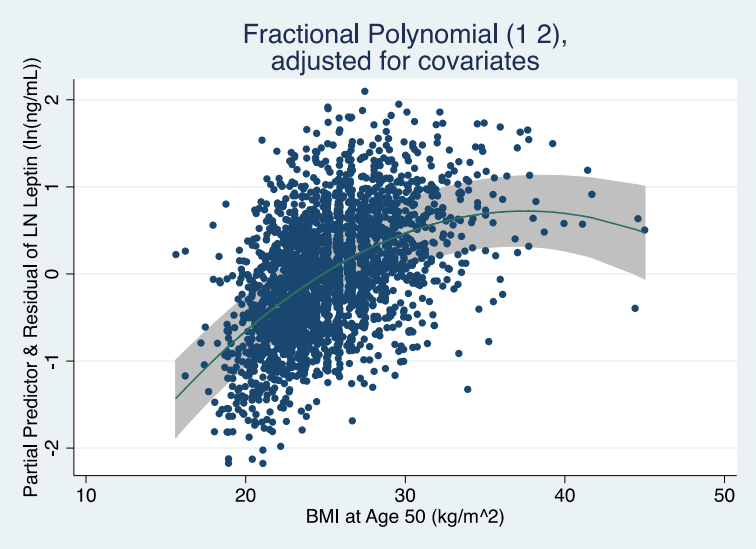

Figure 2. Non-Linear Association between BMI at Age 50 and LN Leptin among Ever-Smokers

The associations of metabolic markers on lung cancer (model one) are summarized in Table S9. The effect of BMI and dietary factors were adjusted in model two and three respectively. Higher LN C-peptide and LN hsCRP were associated with an increased odds ratio for lung cancer after controlling for education, smoking intensity, and smoking duration. Higher LN leptin was associated with a lower odds ratio. Comparing model two with model one in Table S9, although the effect estimate of the additional explanatory variable BMI was insignificant (P = 0.118), it shrank the magnitude of effect estimates of leptin by 23.6% (OR from 0.746 to 0.806). The inclusion of BMI in model two reduced the effect size of leptin on lung cancer. The relationship between leptin and lung cancer could be partly explained by BMI. A comparison between model one and model three demonstrated the additional effects of dietary factors: LN fruits and vegetables daily frequency and supplemental beta-carotene intake. Both dietary factors were significantly associated with lower odds of lung cancer along with metabolic markers after adjusting for covariates. Adding dietary factors increased effect magnitudes of C-peptide and leptin by 23.9% (OR = 1.205 versus 1.254) and 16.5% (OR = 0.746 versus 0.704), respectively. C-peptide and leptin

were mediators that explain the associations between dietary factors and lung cancer. Overall, model three fitted the association better than model one (McFadden's pseudo-R^2 = 0.221 versus 0.196).

In Table S9, model two and three have separately evaluated the relationships between BMI, dietary factors, and lung cancer along with metabolic markers. The fully adjusted model in Table 9 summarizes the effect estimates and P-values of all explanatory variables in comparison with the model three from Table S9. Out of 2245 complete observations, 2002 individuals were included in the final model after excluding 243 concordant observations. There was 22.1% improvement in the final model over a null model which solely contains an intercept (McFadden's pseudo-R^2 = 0.221).

In the final model, the effects of educational achievement, smoking intensity, and smoking duration were adjusted. Each one unit (kg/m^2)increase in BMI at age 50 reduced the lung cancer odds ratio by 3.2% at the significance level of 0.1 (OR = 0.968; 95% CI, 0.931-1.006; *P* = 0.100). Each one unit increase in LN fruits and vegetables daily frequency lowered the odds ratio by 10.9% (OR = 0.891; 95% CI, 0.811-0.979; *P* = 0.016). Odds of lung cancer decreased 14.1% with each 1000 mcg increased in daily supplemental beta-carotene intake (OR = 0.859; 95% CI, 0.763-0.966; *P* = 0.012). The non-linear association between smoking intensity (number of cigarettes/day) and probability of lung cancer is illustrated in Figure 3. Each one unit (ln(pmol/L)) increase in C-peptide concentration increased odd by 28.1% (OR = 1.281; 95% CI, 1.072-1.532; *P* = 0.007). For each one unit (ln(mg/L)) increase in hsCRP, the odds increased by 20.6% (OR = 1.206; 95%, 1.075-1.354; *P* = 0.001). Lastly, each one (ln(ng/Ml)) increase in LN leptin was associated with 23.1% lower odd (OR = 0.769; 95% CI, 0.633-0.934; *P* = 0.008). In summary, fruits and vegetables daily frequency and supplemental beta-carotene intake were protective factors against lung cancer in adjusted analysis. The selected metabolic markers showed significant associations with lung cancer. The approaching significance of BMI at age 50 might be in part explained by the joint action of dietary factors and metabolic markers.

Collinearity is evaluated in the final model with variance inflation factors (VIFs) and tolerance values (Table 10). All VIFs were smaller than two with a mean VIF of 1.12. Tolerances were substantially greater than 0.1. The smallest tolerance value was 0.794. Therefore, collinearity was not a major concern in the final model.

Table 9. Final Multivariable Conditional Logistic Regression for Dietary Factors, Body Mass Index (BMI), and Metabolic Markers Associated with Lung Cancer in 2002 Ever-Smokers

	Adjusted Model†		Fully Adjusted Model‡	
Explanatory Variables	**Odds Ratio (95% CI)**	***P***	**Odds Ratio (95% CI)**	***P***
LN Fruits&Vegetables, per 1 ln(daily frequency) increase	0.893 (0.813–0.980)	0.017	0.891 (0.811–0.979)	0.016
Supplemental beta-carotene, per 1000 mcg/day increase	0.859 (0.764–0.967)	0.012	0.859 (0.763–0.966)	0.012
BMI at age 50, per 1 kg/m^2 increase	-	-	**0.968 (0.931–1.006)**	**0.100**
LN CP, per 1 ln(pmol/L) increase	1.254 (1.050–1.495)	0.012	1.281 (1.072–1.532)	0.007
LN hsCRP, per 1 ln(mg/L) increase	1.212 (1.080–1.359)	0.001	1.206 (1.075–1.354)	0.001
LN Leptin, per 1 ln(ng/Ml) increase	**0.704 (0.592–0.837)**	**<0.001**	**0.769 (0.633–0.934)**	**0.008**
$n_{control}$, n_{case}; AIC, BIC	1290, 747; 1165.690, 1210.644		1262, 740; 1146.819, 1197.236	
P of overall model, pseudo-R^2	<0.001, 0.221		<0.001, 0.221	

†The model adjusting for education, smoking intensity, and smoking duration

‡The final model adjusting for education, smoking intensity, and smoking duration

A non-linear association between smoking intensity and lung cancer odds could not be interpreted directly. For an illustration of the non-linearity, see Figure 3.

The transformation formula: (number of cigarettes per day/10)^-1

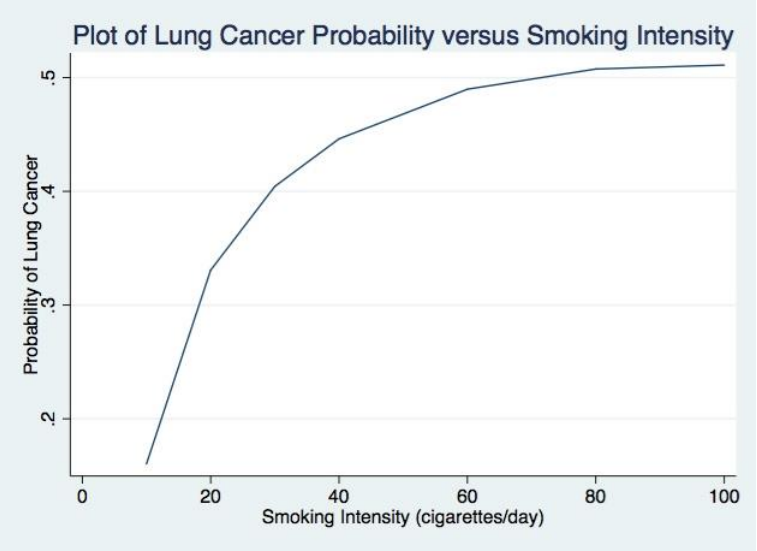

Figure 3. Non-Linear Association between Smoking Intensity and Probability of Lung Cancer among Ever-Smokers

Table 10. Collinearity Evaluation for the Final Model

Explanatory Variables	Variance Inflation Factors (VIF)	Tolerance
Education, per 1 level increase	1.04	0.959
BMI at age 50, per 1 kg/m^2 increase	1.15	0.869
LN Fruits&Vegetables, per 1 ln(daily frequency) increase	1.11	0.903
Supplemental beta-carotene, per 1000 mcg/day increase	1.02	0.976
Transformed smoking intensity with power of -1, per 1 cigarette/day increase	1.09	0.916
Smoking duration, per 1-year increase	1.16	0.859
LN CP, per 1 ln(pmol/L) increase	1.13	0.887
LN hsCRP, per 1 ln(mg/L) increase	1.11	0.903
LN Leptin, per 1 ln(ng/Ml) increase	1.26	0.794
Mean VIF	1.12	

4.3 Assumption Checking

1) Independence of Errors

In the study, observations were independent of each other. No duplicated responses occurred in the data.

2) Linearity

Locally weighted regression between numeric variables was visually examined by performing Lowess smoothing plots and spline modeling. Skewed distributed explanatory variables were transformed into a natural log form except for supplemental beta-carotene intake whose transformation was not helpful. For all linear and conditional logistic regressions, possible non-linear associations were considered by conducting MFP analysis.

3) Lack of Influential Outliers

As shown in Figure 4 and 5, influential outliers were inspected by plotting Pearson standardized residual and Cook's distance against observation number for the final model. In Figure 4, the majority of the information points are located in the range of -1.96 and 1.96 on the y-axis. Cook's distances in Figure 5 range from 0 and 0.15, thus influential outliers were not a cause of concern in the final model.

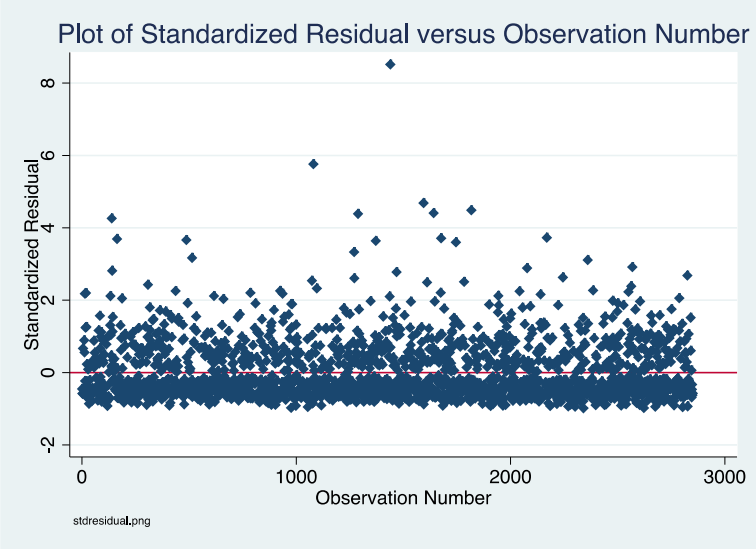

Figure 4. Inspection of Influential Observations with Pearson Standardized Residual in the Final Model

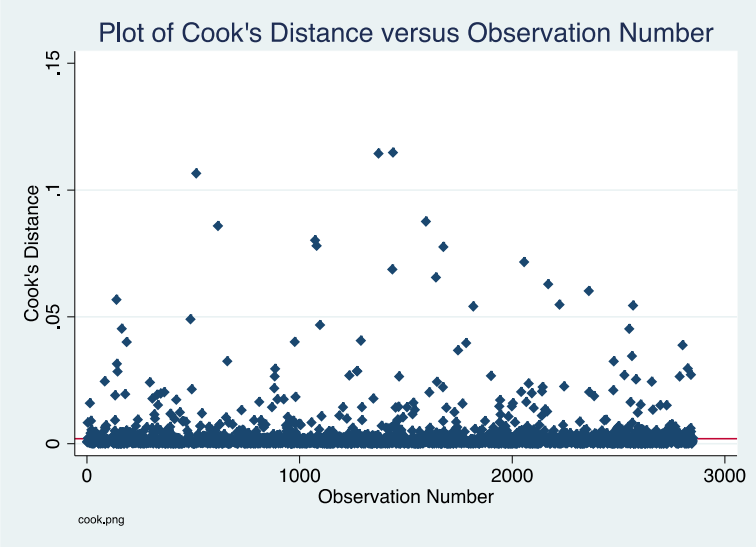

Figure 5. Inspection of Influential Observations with Cook's Distance in the Final Model

4.4 Model Diagnostics

1) Measures of Fit

Akaike information criterion (AIC), Bayesian information criterion (BIC), and R^2 or pseudo-R^2 of intermediate analyses were evaluated to find optimal models. The final model has an AIC of 1146.819, a BIC of 1197.236, and a pseudo-R^2 of 0.221 (Table 9).

2) Model Adequacy/Specification Error

The model specification was evaluated by conducting a link test. The linear predicted value (_hat) is statistically significant with a P-value of <0.001. The P-value of the linear predicted value squared (_hatsq) is 0.920. Therefore, the model was appropriately specified as the link test was insignificant.

4.5 Exploratory Analyses

The final model was applied to the subgroups of lung cancer types: non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC) with their matched group of control. Then in NSCLC, the comparison was drawn between adenocarcinoma, squamous cell carcinoma, and Other NSCLC subgroups.

The analysis of the NSCLC group included 973 controls and 736 cases (Table 11). The overall model was significant with a P-value of <0.001 and a pseudo-R^2 of 0.225. Except for educational achievement and BMI at age 50, all explanatory variables were significant. The analysis of SCLC was comprised of 174 controls and 119 cases (Table 11). The overall model was significant with a P-value of <0.001 and a pseudo-R^2 of 0.225. Comparing the effect estimates of NSCLC analysis to those of SCLC analysis demonstrated similarities and differences between the two groups regarding study explanatory factors. The protective effect of BMI exists in NSCLC but not SCLC. Odds ratios for fruits and vegetables daily frequency and supplemental beta-carotene intake were less than one in both lung cancer subgroups. The effect size of LN fruits and vegetables daily frequency was larger in NSCLC than in SCLC.

The effect estimates of C-peptide and hsCRP were greater than one in both NSCLC and SCLC. The effect magnitude of C-peptide was 6.8 times more prominent in NSCLC than in SCLC. The odds ratios for leptin were less than one in both groups. The effect of leptin on NSCLC was 148% higher than the effect in SCLC. In other words, the overall model fitted NSCLC and SCLC in a similar way, but the strengths of associations for individual explanatory variables were different. In this study, dietary factors, BMI, and metabolic markers fitted NSCLC group better than SCLC group.

Among the subgroups of NSCLC, 432 controls and 319 cases of adenocarcinoma were included (Table 12). The overall model was significant with a P-value of <0.001 and a pseudo-R^2 of 0.192. In the analysis of squamous cell carcinoma, 245 controls and 188 cases were studied. The model for squamous cell carcinoma was significant with a P-value of <0.001 (pseudo-R^2 = 0.324). The analysis of Other NSCLC consisted of 296 controls and 229 cases. The association with Other NSCLC was significant (P < 0.001; pseudo-R^2 = 0.283). The explanatory variables fitted the squamous cell carcinoma group better than adenocarcinoma and Other NSCLC groups. The effect of BMI was protective in all three subgroups, and the effect size in the Other NSCLC subgroup was the largest. The protective effect of fruits and vegetables was most prominent in Other NSCLC and was close to the null and non-significant for adenocarcinoma and squamous cell carcinoma. The effect estimates of C-peptide were inconsistent across the three subgroups. The positive association between C-peptide and lung cancer appeared to be absent in squamous cell carcinoma while present in adenocarcinoma and Other NSCLC. HsCRP was associated with higher odds for all NCSLC types with the strongest effect found in the Other NSCLC subgroup. The inverse association between leptin and lung cancer were observed in all three NSCLC groups.

Table 11. Multivariable Conditional Logistic Regression for Dietary Factors, Body Mass Index (BMI), and Metabolic Markers Associated with Non-Small Cell Lung Cancer (NSCLC) and Small Cell Lung Cancer (SCLC)

$n_{control}$, n_{case} **Explanatory Variables**	**NSCLC (973, 736)** **Odds Ratio (95% CI), *P***	**SCLC (174, 119)** **Odds Ratio (95% CI), *P***
Education, per 1 level increase	0.937 (0.872–1.006), 0.073	0.877 (0.719–1.069), 0.194
Smoking intensity, per 1 cigarette/day increase	N/A†, <0.001	1.017 (0.992–1.042), 0.183
Smoking duration, per 1-year increase	1.064 (1.050–1.079), <0.001	1.092 (1.045–1.140), <0.001
LN Fruits&Vegetables, per 1 ln(daily frequency) increase	0.879 (0.794–0.974), 0.014	0.939 (0.727–1.212), 0.628
Supplemental beta-carotene, per 1000 mcg/day increase	0.868 (0.764–0.985), 0.028	0.809 (0.565–1.159), 0.248
BMI at age 50, per 1 kg/m^2 increase	0.963 (0.923–1.005), 0.084	**1.000 (0.904–1.108), 0.992**
LN CP, per 1 ln(pmol/L) increase	1.340 (1.102–1.629), 0.003	**1.050 (0.656–1.681), 0.838**
LN hsCRP, per 1 ln(mg/L) increase	1.208 (1.067–1.367), 0.003	1.231 (0.881–1.719), 0.224
LN Leptin, per 1 ln(ng/Ml) increase	0.750 (0.607–0.927), 0.008	0.899 (0.544–1.484), 0.676
n; AIC, BIC	n = 1709; 976.574, 1025.567	n = 293; 182.749, 215.871
P of overall model, pseudo-R^2	<0.001, 0.225	<0.001, 0.225

† A non-linear association between smoking intensity and odds of NSCLC could not be interpreted directly.
The transformation formula: (number of cigarettes per day/10)^-1

Table 12. Multivariable Conditional Logistic Regression for Dietary Factors, Body Mass Index (BMI), and Metabolic Markers Associated with Adenocarcinoma, Squamous Cell Carcinoma, and Other Non-Small Cell Lung Cancer (NSCLC)

$n_{control}$, n_{case} **Explanatory Variables**	**Adenocarcinoma (432, 319)** **Odds Ratio (95% CI), *P***	**Squamous cell carcinoma (245, 188)** **Odds Ratio (95% CI), *P***	**Other NSCLC (296, 229) †** **Odds Ratio (95% CI), *P***
Education, per 1 level increase	0.894 (0.800–0.998), 0.046	0.986 (0.845–1.149), 0.853	0.925 (0.810–1.057), 0.253
Smoking intensity, per 1 cigarette/day increase	N/A‡, <0.001	1.012 (0.995–1.030), 0.163	N/A§, <0.001
Smoking duration, per 1-year increase	1.048 (1.029–1.068), <0.001	1.135 (1.088–1.183), <0.001	1.062 (1.036–1.088), <0.001
LN Fruits&Vegetables, per 1 ln(daily frequency) increase	0.941 (0.813–1.088), 0.409	0.937 (0.764–1.148), 0.531	0.753 (0.602–0.941), 0.013
Supplemental beta-carotene, per 1000 mcg/day increase	0.943 (0.770–1.154), 0.567	**0.706 (0.532–0.937), 0.016**	0.928 (0.754–1.142), 0.480
BMI at age 50, per 1 kg/m^2 increase	0.970 (0.913–1.030), 0.318	0.987 (0.899–1.085), 0.790	0.947 (0.869–1.032), 0.216
LN CP, per 1 ln(pmol/L) increase	1.502 (1.133–1.992), 0.005	**0.908 (0.577–1.429), 0.677**	1.535 (1.044–2.258), 0.029
LN hsCRP, per 1 ln(mg/L) increase	1.161 (0.967–1.393), 0.109	1.166 (0.879–1.546), 0.288	1.432 (1.128–1.817), 0.003
LN Leptin, per 1 ln(ng/Ml) increase	0.770 (0.559–1.060), 0.108	0.785 (0.505–1.222), 0.284	**0.583 (0.387–0.879), 0.010**
n; AIC, BIC	n = 751; 457.437, 499.030	n = 433; 229.669, 266.306	n = 525; 290.129, 328.499
P of overall model, pseudo-R^2	<0.001, 0.192	<0.001, 0.324	<0.001, 0.283

† Other NSCLC in this study includes large cell carcinoma, other NSCLC, not otherwise specified carcinoma, and other/missing.
‡ A non-linear association between smoking intensity and odds of Adenocarcinoma
The transformation formula: (number of cigarettes per day/10)^-1
§ A non-linear association between smoking intensity and odds of Other NSCLC
The transformation formula: (number of cigarettes per day/10)^-2

CHAPTER 5 DISCUSSION

5.1 Explanatory Variables of Lung Cancer

This study identified that an increased risk of lung cancer was associated with less daily frequency of fruits and vegetables, less supplemental beta-carotene intake, lower BMI at age 50, and higher C-peptide, higher hsCRP, and lower leptin serum concentrations in ever-smokers. Higher fruits and vegetables intake and higher supplemental beta-carotene intake protected people against lung cancer, but meanwhile, their protective effects were partly offset by the effect of decreasing BMI. The beneficial effect of higher fruits and vegetables intake was partly offset by increasing C-peptide. Lowering hsCRP was another effect of fruits and vegetables intake on lowering lung cancer risk. Although a complete understanding of the process in the pathophysiology of lung cancer was not attained, the findings of this study help improve our understanding of lung carcinogenesis and may lead to public health approaches to improve lung cancer prevention.

5.1.1 Relationships between Dietary Factors and Lung Cancer

Some of the dietary factors observed to be important in the univariate analysis were eliminated in multivariable analysis as several factors were correlated with each other and presented similar attributes of individuals' diets. For example, higher daily frequency of fruits and vegetables intake was accompanied by higher vitamin C intake from diet and higher total isoflavone intake, therefore relationships between the latter factors and lung cancer were suppressed in multivariable models. As the odds ratio was 0.891 for each one unit increase in daily frequency of fruits and vegetables, the odds of lung cancer decreased by 10.9% (95% CI, 0.811-0.979; P = 0.016). The finding of the current study aligned with a 14% risk reduction in lung cancer reported in an earlier study (Vieira et al., 2016).

The relationship between supplemental beta-carotene and lung cancer remained unclear. In the current study, we found that higher daily supplemental beta-carotene intake was associated with lower odds of lung cancer (OR = 0.859; 95% CI, 0.763-0.966; P = 0.012). In contrast, previous studies

demonstrated a detrimental effect or no effect of daily supplemental beta-carotene on lung cancer risk. Participants who received 20 mg (133,333 IU) of synthetic beta-carotene per day had 6% higher lung cancer risk than the placebo group in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention (ATBC) Study (Virtamo et al., 2003). In the Carotene and Retinol Efficacy Trial (CARET), lung cancer risk in the intervention group with a daily combination of beta-carotene (30 mg = 15,000 mcg RAE) and retinol (25,000 IU = 7500 mcg RAE) was 12% higher than the risk in control group (Goodman et al., 2004). In another randomized control trial, no significant difference was found in lung cancer risk between the group with 50 mg (25,000 mcg RAE) beta-carotene every other day and the placebo group (RR = 1.00; 95% CI, 0.85-1.17) (Lin et al., 2009). It is worth noting that the three studies were experimental studies and the study groups received dosages substantially higher than the recommended level. When converted to retinol activity equivalents (RAE), the measure for beta-carotene reflects the active amount of provitamin A carotenoids from food to retinol or active vitamin A (Trumbo, Yates, Schlicker, & Poos, 2001). One RAE equals to 12 mcg beta-carotene from foods or 2 mcg beta-carotene from supplements. In the current study, the average daily dietary and supplemental beta-carotene intake were 4,820.5 mcg/day (651.7 mcg RAE) and 475.7 mcg/day (237.9 mcg RAE). In other words, a moderate dosage of supplemental beta-carotene does not appear to be harmful. On the other hand, the association between beta-carotene supplement and lung cancer was adjusted for other study dietary factors, namely BMI, metabolic markers, and cigarette smoking. The effect of supplemental beta-carotene is less likely to be due to residual confounding in the current study.

5.1.2 Relationships between Dietary Factors, BMI, and Lung Cancer

The current study found a protective effect of increasing BMI at age 50 in ever-smokers with an odds ratio of 0.968, corresponding to a 3.2% risk reduction per one unit (kg/m^2) increase (95% CI, 0.931-1.006; *P* = 0.100). In a systematic review by Duan et al. (2015), investigators found that each five units increase in BMI was associated with a 3.3% decrease in the risk of lung cancer. The effect size in the

current study was stronger (OR = 0.850 per five units increase in BMI) than the findings in Duan et al.'s study (2015). This may be attributed to a more diverse participant pool studied. Including the group of never-smoker might have attenuated the effect size of BMI. Unlike the linear association we found in the current study, the relationship between BMI and lung cancer risk was non-linear in the study by Duan et al. (2015) They also conducted a dose-response analysis to quantify the effect of BMI on lung cancer risk in different BMI categories. With normal weight (18.5 - 24.9 kg/m^2) as the reference category, they found the protective effect of higher BMI among people with BMI of 35 kg/m^2 (summary RR = 0.81; 95% CI: 0.72-0.91), 30 kg/m^2 (summary RR = 0.91; 95% CI: 0.85-0.98), and 25 kg/m^2 (summary RR = 0.98; 95% CI: 0.95-1.01) (Duan et al., 2015). This further confirmed the inverse association between BMI and lung cancer risk. In a case-control study by Sanikini et al. (2018), BMI were classified into four categories: underweight (less than 18.5 kg/m^2), normal weight (18.5 - 24.9 kg/m^2), overweight (25.0 - 29.9 kg/m^2) and obese (30 kg/m^2 or greater). They evaluated the association between ordinal weight category and lung cancer risk. Lower lung cancer risks were detected for people in overweight category (OR = 0.77; 95% CI: 0.68-0.86) and in obese category (OR = 0.69; 95% CI: 0.59-0.82) in comparison with normal weight category after adjusting for age, sex, study center, time elapsed, pack-years of smoking, and education level (Sanikini et al., 2018). To further examine the relationship between BMI and lung cancer risk among subgroups, investigators stratified observations by smoking status. Compared to people with normal weight, overweight and obese participants were at 30% and 45% lower risk of lung cancer, respectively in former-smokers. The effect sizes were smaller in current-smokers: 21% and 25% risk reductions (Sanikini et al., 2018). As there was adjustment for smoking duration and intensity in the final model for this study, the BMI effect genuinely existed in the association with lung cancer independent of smoking exposure.

The relationship between dietary factors and BMI found that fruits and vegetables daily frequency was inversely associated with BMI (β = -0.407; 95% CI, -0.660- -0.155). In a meta-analysis of

11 RCTs and six prospective cohort studies by Hebden et al. (2017), long-term weight gain in adults was explained by lower consumption of whole fruits. The beneficial effect of higher whole fruits consumption on body weight management was mediated by smaller total energy intake (Hebden et al., 2017). In the current study, a positive association was found between total energy from diet and BMI (β = 0.600; 95% CI, 0.137-1.064).

Previous studies have found an inverse association between beta-carotene and excess body fat in plasma and adipocyte. A lower plasma beta-carotene concentration was related to a higher BMI in one study (Brady, Mares-Perlman, Bowen, & Stacewicz-Sapuntzakis, 1996). The actual concentration of beta-carotene in the adipocyte was inversely related to the measure of obesity (Östh, Öst, Kjolhede, & Strålfors, 2014). Beta-carotene concentrations in the adipocytes from both obese non-diabetic and obese diabetic participants (BMI ≥ 28 kg/m^2) were substantially lower than that from the lean (BMI < 23 kg/m^2) or non-obese participants (23 ≤ BMI < 28 kg/m^2) (Östh et al., 2014). The association with BMI is unclear as the evidence for moderate supplemental beta-carotene intake is scarce in observational studies. In the current study, a non-linear association was suggested by the multivariable fractional polynomial (MFP) model: supplemental beta-carotene was associated with BMI (Table S5). Regardless of techniques for measurements, lower beta-carotene levels in supplemental intake, plasma, and adipocyte were associated with a higher BMI.

5.1.3 Relationships between Dietary Factors, BMI, Metabolic Markers, and Lung Cancer

Red meat consumption (β = 0.053; 95% CI, 0.024-0.083; *P* < 0.001) and fruits and vegetables daily frequency (β = 0.030; 95% CI, -0.001- 0.060; *P* = 0.057) were positively associated with C-peptide concentration in the current study (Table S7a). The positive association between red meat consumption and C-peptide concentration concurred with previous studies. The work by Fung et al. (2001) found that Western dietary patterns, a key component is higher red meat consumption, that in turn was positively correlated with fasting insulin and C-peptide level (Fung et al., 2001). The positive relationships between

the Western diet, serum C-peptide, and serum insulin were also recognized in a previous study (Kerver, Yang, Bianchi, & Song, 2003). However, the role of fruits and vegetables in the association with C-peptide level we found does not align with these studies. The Prudent dietary pattern characterized by high fruits and vegetables intake was inversely correlated with insulin (Fung et al., 2001). Higher fruits or green leafy vegetables intake was associated with a significantly reduced risk of type 2 diabetes (Li, Fan, Zhang, Hou, & Tang, 2014). The reasons behind the contradictory observation in the current study related to the different definition of fruits and vegetables and the measure of dietary factors. Categorization of fruits and vegetables in the current study included juice, excluded fried potatoes, and adjusted for ketchup, onion, and garlic amounts. The food types and level of preparation such as whole or processes foods may also play a role in metabolic activities. The metabolic responses to processed juices, which often contain high quantities of sugar differ from response to whole fruits and vegetables. On the other hand, fruits and vegetables measured in daily frequency might not accurately reflect absolute intake.

A higher daily frequency of fruits and vegetables (β = -0.049; 95% CI, -0.084- -0.014; *P* = 0.007) and a lower red meat consumption (β = 0.113; 95% CI, 0.060-0.167; *P* < 0.001) were related to a lower hsCRP concentration, which indicates a Prudent diet rich in fruits and vegetables helps prevents inflammation (Table S7b). The anti-inflammatory effect might be attributable to salicylates, vitamins, and minerals in whole plants (Smidowicz & Regula, 2015). In red meat preparation, people may be exposed to polycyclic aromatic hydrocarbons via smoking or grilling meat at high temperatures. As well excess haem iron from red meat may induce inflammatory responses by mediating hsCRP (World Cancer Research Fund/American Institute for Cancer Research [WCRF/AICR], 2018).

Increasing leptin suppresses appetite to maintain energy balance. In the associations between dietary factors and leptin, we found that higher red meat intake (β = 0.078, 95% CI, 0.044-0.111; *P* < 0.001) and vitamin C intake from diet (β = 0.101; 95% CI, 0.008-0.193; *P* = 0.033) were associated with a

higher leptin level (Table S7c). In the previous study by Ko et al. (2016), they also found a diet with higher Western diet scores was associated with a higher leptin concentration. The inverse association between daily frequency of fruits and vegetables (β = -0.051; 95% CI, -0.102- -0.001; *P* = 0.047) and leptin concentration also aligns with Ko et al.'s (2016) finding of a Prudent diet. But the positive association between vitamin C intake from diet and leptin is contrary to expectations. Because vitamin C and fruits and vegetables daily frequency were highly correlated, they were expected to relate to leptin level in a similar way. A previous study by Aeberli et al. (2006) on the effects of dietary intake on inflammatory markers found that both vitamin C intake and beta-carotene intakes were negative explanatory variables of leptin concentration.

In the experimental study by Jamalan, Rezazadeh, Zeinali, and Ghaffari (2015), participants' mean leptin level was significantly lower (mean = 19; SD = 12) after receiving vitamin C 1000 mg/day for four weeks than at baseline (mean = 31; SD = 15 ng/mL) (Jamalan, Rezazadeh, Zeinali, & Ghaffari, 2015). In Table 6, vitamin C intake from diet demonstrated a protective effect against lung cancer in the unadjusted univariate analysis. But no association between vitamin C and lung cancer was found in the fully adjusted model (Table 9). Vitamin C had no directly independent effect on lung cancer. However, increasing BMI and increasing leptin caused by higher vitamin C were associated with a lower risk of lung cancer. In the current study, a non-linear association existed between supplemental beta-carotene intake and leptin concentration (Table S7c). It seems that reduced leptin associated with higher supplemental beta-carotene stimulates appetite and energy storage.

In the study by Abdullah, Hasan, Raigangar, and Bani-Issa (2012), a higher C-peptide in the over-weight group than the non-overweight group indicated that C-peptide, a measure of endogenous insulin secretion, was positively associated with BMI (Abdullah, Hasan, Raigangar, & Bani-Issa, 2012). In our study, we detected a non-linear association between C-peptide and BMI (Table S8a). However, the

effects of C-peptide and BMI associated with lung cancer were contrary against each other. The risk reduction in lung cancer by increasing BMI was stronger than the risk imposed by increasing C-peptide.

We found a higher BMI was associated with a higher hsCRP level (Table S8b), which is in agreement with previous studies (β = 0.021; 95% CI, 0.006-0.035; *P* = 0.005) (Clark et al., 2016; Visser M et al., 1999). As mentioned before, BMI and lung cancer risk were inversely associated. The effect of increasing BMI on lowering lung cancer risk had offset the effect of increasing hsCRP. The hsCRP concentrations in female participants were significantly higher compared to male participants in the current study. In addition, sex interacted with the effect of BMI on hsCRP ($P_{interaction}$ = 0.005). In the study of Clark et al. (2016), they also found the association between BMI and hsCRP was stronger in female individuals than in male individuals. This might be due to a few biological reasons: different metabolic activity of adipose tissue in males versus females; higher leptin concentrations in female participants; and higher body fat and different fat distribution in female participants (Choi, Joseph, & Pilote, 2013).

In the current study, a non-linear relationship was detected between BMI at age 50 and leptin level (Table S8c and Figure 2). However, the non-linear association may have resulted from a few extreme values among people with BMI greater than 40 kg/m^2. The concentration of leptin in female participants was higher than in male participants. A previous study by Kazmi et al. found similar results. The serum leptin level was lower in the non-obese group than the group with excess body fat. Female participants had higher leptin levels than male participants in both groups (Kazmi et al., 2013). However, a greater age was associated with a higher leptin concentration, which was not found in the study by Kazmi et al. A relatively small study sample with a younger age range (25 to 40 old) might have prevented them from finding the effect of age in relation to leptin concentration.

The current study found that C-peptide concentration was positively associated with lung cancer (adjusted OR = 1.281; 95% CI, 1.072-1.532; *P* = 0.007). Such an association was independent of BMI. Findings of the current study concurred with previous studies. In the study by Hsu et al. (2013), for every

one unit (pmol/mL) increase in serum C-peptide concentration, there was an associated 165% increased risk of lung cancer death in female participants (95% CI, 1.31-5.36). In our study, sex did not play a significant role in the risk of lung cancer. This suggested that increasing C-peptide might help predict lung cancer risk and guide screening in both sexes. According to the study conducted by Argirion, Weinstein, Männistö, Albanes, and Mondul (2017), the mean fasting insulin concentration in lung cancer cases was 8.7% higher than the mean in controls. Males participants in the highest quartile of insulin level were at 110% higher risk of lung cancer in comparison with people in the lowest quartile (HR = 2.10; 95% CI, 1.12-3.94) (Argirion et al., 2017). As the current study utilized data that collected non-fasting serum samples, the association between insulin and lung cancer could not be evaluated. An increase in either C-peptide or insulin concentration indicates insulin resistance, a precursor of type 2 diabetes mellitus. Individuals taking metformin to control diabetes may have an advantage in terms of reduced risk of lung cancer. These finding may suggest that conventional intervention for diabetes-related diseases, like metformin, may help inhibit lung cancer initiation and progression (Argirion et al., 2017; Zhang et al., 2014).

hsCRP level was found to be positively associated with lung cancer (adjusted OR = 1.206; 95%, 1.075-1.354; *P* = 0.001). A previous nested case-control study using data from the PLCO trial found similar results: the combination of interleukin 8 and C-reactive protein (CRP) were robust metabolic markers in predicting lung cancer (Pine et al., 2011). Another nested case-control study in the PLCO trial matched participants on age, sex, entry year, follow-up time, and smoking (pack-year and quit-time) (Chaturvedi et al., 2010). Increased CRP levels were associated with higher lung cancer risk in former-smokers and current-smokers, but not in never-smokers. In current-smokers, the total number of cigarettes they had smoked (pack-year) was associated with CRP level. Therefore, the concentration of CRP, an indicator of inflammation, might just reflect how the body responded to cigarette-smoking (Agassandian, Shurin, Ma, & Shurin, 2014; Chaturvedi et al., 2010). However, in the current study, the

association between hsCRP and lung cancer exists after adjusting for smoking exposures (intensity and duration). The level of hsCRP was not only a response to smoking but also a mediator in lung cancer development.

In a previous PLCO study on metabolic markers, the inverse association between leptin and lung cancer had been attenuated after adjusting for BMI (Goodwin et al., 2015). In the current study, lung cancer odds decreased by 23.1% as leptin concentration increased one unit (ln(ng/mL)) (95% CI, 0.633-0.934; *P* = 0.008) after adjusting for other explanatory variables and covariates, including BMI. In Table S9 Model 3, the effect estimate of leptin without adjustment for BMI was 0.704 (95% CI, 0.592-0.837, *P* < 0.001). After adding BMI in the analysis, the protective effect of leptin decreased by 22.0%. Therefore, higher BMI in part explained the effect of increasing leptin on lowering lung cancer risk. In addition, the effect of leptin on lung cancer was prominent and independent of BMI.

5.2 Limitations

The study matched controls to cases on factors associated with study exposures using PLCO data. This might have introduced selection bias as the sample association of exposure with lung cancer was altered by the matched study design (Rothman, Greenland, & Lash, 2008). Because controls were matched to cases on sex, age, and race/ethnicity, the associations of lung cancer with regard to these matching factors could not be evaluated. Also, as the matching factors were associated with study exposures, the consequential overmatching may have harmed statistical efficiency (Rothman, Greenland, & Lash, 2008). If the onset of lung cancer occurred before metabolic markers were collected, the temporal relation could not be determined. However, this study was nested in the intervention cohort of the prospective PLCO study, hence the possibility of reverse causation bias is low. Besides, the results obtained in our research could be influenced by several limitations: inaccessible information, self-reported data, and restricted population representation.

5.2.1 Unattainable Information

Only BMI was included to represent participants' excess adiposity. As an indirect method to measure obesity, BMI has limitations (Egom et al., 2018). Higher BMI simply indicates higher body mass; this could mean greater body fat and muscle mass. BMI alone could not describe a person's adiposity profile. Hence, a full picture of obesity status could not be obtained in the study. The current study was based on an assumption that participants' diet did not change over time. However, people might adjust diet according to changing preferences, food availability, and health concerns. The dietary information could not capture potential changes over time.

5.2.2 Self-reported Data

The information on baseline characteristics and dietary factors were gathered from participant-reported questionnaires (Prorok et al., 2000). In the baseline form, respondents provided their BMI, health history, and smoking history. To correspond with ideal societal norms, participants might over-report favourable attributes in diet. On the other hand, individuals may have been more likely to minimize undesirable behaviours in cigarette smoking. Thus, information on these factors might have been affected by social desirability bias. If cigarette smoking exposures were underreported, the harmful effect might shift toward a weaker relation than in reality. The effect magnitude of fruits and vegetables daily frequency might be underestimated as well. Another non-differential misclassification bias is caused by memory error in dietary information. Because the occurrences of this error were not thought to be determined by lung cancer status, the error might weaken the effect estimates for dietary factors. Because measures in the baseline and dietary were collected at one point in time, reliability does not apply (Patten, 2015).

5.2.3 Restricted Population Representation

The study sample was drawn from a more educated population with less current-smokers and fewer people with obesity (Pinsky et al., 2007). Given the age range of 55 to 74 years at study entry,

participants were at an advanced age and at a higher risk of having lung cancer (Prorok et al., 2000). In this regard, individuals who volunteered to participate in the PLCO screening trial were more likely to be healthier than the overall population in the United States. The "healthy volunteer effect" might introduce selection bias and weaken external generalizability of findings (Pinsky et al., 2007). As a result, the inferences drawn from the study might have difficulty in generalizing to a wider population. Although the findings were based on data from the United States, the associations identified are likely to be applicable to ever-smoker populations in different settings because they are thought to result from universal biological mechanism in humans. The relationships between dietary factors, BMI, metabolic markers, and lung cancer were only described in the high-risk group (ever-smokers). The group of never-smokers was excluded in this study, so it is unreasonable to generalize the findings to never-smokers.

5.3 Strengths

One of the main strengths of the current study was the prospective study design. The matched case-control study nested in the PLCO trial was less likely to be affected by disease-differential bias. Assessment of dietary intake at study baseline should not have been influenced by lung cancer related or precursor conditions-associated diet modifications. The information on body composition had been assessed using BMI at age 50. Serum samples were collected at least more than one year before lung cancer diagnosis. Therefore, the study was not susceptible to recall bias. Cases and controls were both from the same explicitly described PLCO trial cohort, this minimized one of the commonest vulnerabilities in case-control design—selection bias (Sedgwick, 2014). Besides, the study was endowed with 2 advantages of the conventional case-control design. First, the study was efficient at evaluating a chronic disease—lung cancer—that usually has a long latency between exposures and disease manifestation (Howard, 2015). Second, the design was cost-effective and allowed us to examine

multiple exposures simultaneously, for instance, a series of dietary factors and metabolic markers (Patten, 2015).

At the analysis stage, robust statistical methods were utilized to improve statistical precision. The study had adjusted for interactions and covariates. For all logistic regressions evaluating the risk of lung cancer, conditional analyses had controlled for confounding by the matching and by inclusion of potential confounders as explanatory variables in the models (Pearce, 2016). Multivariable linear regression that assesses associations between quantitative variables was performed with cluster-robust standard errors to adjust for the matched data. Meanwhile, non-linear associations between quantitative variables and lung cancer were examined by employing multivariable fractional polynomial (MFP) models. In the process of model building, we used a liberal P-value cut-point ($P < 0.15$) to purposefully select and retain candidate explanatory variables in intermediate associations (Bursac, Gauss, Williams, & Hosmer, 2008).

Important confounders were carefully controlled in the study. In the association between dietary factors, metabolic markers, and lung cancer, higher BMI had a direct effect on lowering lung cancer. Increasing BMI also had an indirect effect by increasing leptin. This indicated the potentiality of BMI as an important factor to consider for lung cancer risk prediction and screening. Overall, the high-quality data from the PLCO cancer screening trial was helpful for accurately describing the study associations. Carefully following up participants had maximized data completeness. We were able to conduct exploratory analyses in selected lung cancer histological types because of the sufficient sample size.

5.4 Implications

First, the study results provide updated evidence to focus lung cancer prevention scheme on the role of diet and metabolism for ever-smokers from a public health perspective. Underweight smokers are recommended to attain a normal BMI by having a nutritious diet including the nutrients found to be

protective in the current study. Given the evidence of high quality RCT it is prudent to obtain dietary beta-carotene from natural dietary sources, as they also appear to be protective (data not shown). As healthy diet and higher body weight in the normal range are strong and modifiable protective factors for lung cancer in ever-smokers, the potential for preventable incidence and mortality are enormous. To achieve this goal at the population level, the findings can be used to improve health guidelines and educate people, especially high-risk group, to have a good diet and body weight.

Second, the findings of the study may bring new opportunities for improving secondary prevention of lung cancer. Dietary, BMI and metabolic information may improve the performance of current risk prediction model for lung cancer and assist healthcare professionals to detect the disease at an early stage. Adding segments of people's diet, obesity status, and metabolic profile to the lung cancer screening for high risk individuals permits practitioners better identify individuals with a higher likelihood of lung cancer occurrence in Canada.

Third, this study provides new ideas and directions for future efforts or research aimed at the treatment of lung cancer. C-peptide, hsCRP, and leptin are important explanatory variables of lung cancer independent of dietary factors, BMI, and other covariates. The panel of selected metabolic markers may serve as targets for possible future development of chemoprevention and therapeutics. Conventional drugs, for instance metformin used to treat T2DM, act on these metabolic markers, which indicates their undiscovered capabilities to aid in medical interventions for lung cancer (Sleire et al., 2017; Pryor & Cabreiro, 2015).

5.5 Future Directions

In future research, it is crucial to conduct more external validations of our model or identified associations in wider populations with a longitudinal design. Broadening the panel of metabolic markers helps to advance knowledge on associations between insulin metabolism related metabolic markers and lung cancer. Because sex differences exist in the associations of metabolism and lung cancer biology, sex

effects on metabolic markers should be studied and evaluated. If well-defined measures of obesity (a combination of BMI and waist circumference), socioeconomic status, and history of pulmonary disease are available, they can verify and improve identified estimates of association effects. Variable selection with statistical learning had been used to study the role of dietary factors associated with outcome of interest in chronic diseases (Biesbroek et al., 2015; Panaretos et al., 2018). Employing different statistical methods and comparing them with the current study might tell us their capability in the field of cancer epidemiology and further improve our models and understanding of associations.

5.6 Conclusions

In summary, our study contributes to the literature by exploring the associations between dietary factors, BMI, metabolic markers, and the risk of lung cancer and its subgroups. An inverse association was found between higher BMI and the risk of lung cancer. Having a lower BMI at 50 is associated with a higher risk. Higher intake of fruits and vegetables and supplemental beta-carotene in an appropriate dosage range are inversely associated with the risk. Fruits and vegetables and supplemental beta-carotene may act as possible protective factors for lung cancer prevention. The profile of metabolic markers might add value to existing risk prediction model to focus on early detection of lung cancer in high-risk population. An increase in C-peptide and hsCRP and a decrease in leptin were associated with higher lung cancer risk. The selected metabolic markers have the potential to influence intervention planning for chemoprevention or therapeutics if their relationships to lung cancer risk are validated in future studies. More investigations with broad and selected dietary factors and metabolic markers are warranted to validate our findings and explore the factors that contribute to development of lung cancer. Studies with a longitudinal design are needed to verify the etiology of the disease.

REFERENCES

Abdullah, A., Hasan, H., Raigangar, V., & Bani-Issa, W. (2012). C-peptide versus insulin: Relationships with risk biomarkers of cardiovascular disease in metabolic syndrome in young Arab females. *International Journal of Endocrinology, 2012*, https://doi.org/10.1155/2012/420792

Aeberli, I., Molinari, L., Spinas, G., Lehmann, R., l'Allemand, D., & Zimmermann, M. B. (2006). Dietary intakes of fat and antioxidant vitamins are predictors of subclinical inflammation in overweight Swiss children. *The American Journal of Clinical Nutrition, 84*(4), 748–755. https://doi.org/10.1093/ajcn/84.4.748

Agassandian, M., Shurin, G. V., Ma, Y., & Shurin, M. R. (2014). C-reactive protein and lung diseases. *The International Journal of Biochemistry & Cell Biology*, *53*, 77–88. https://doi.org/10.1016/j.biocel.2014.05.016

Alberg, A. J., Brock, M. V., Ford, J. G., Samet, J. M., & Spivack, S. D. (2013). Epidemiology of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest, 143*(5), e1S-e29S. https://doi.org/10.1378/chest.12-2345

Alberg, A. J., Brock, M. V., & Samet, J. M. (2005). Epidemiology of lung cancer: Looking to the future. *Journal of Clinical Oncology, 23*(14), 3175–3185. https://doi.org/10.1200/JCO.2005.10.462

Aleksandrova, K., Mozaffarian, D., & Pischon, T. (2018). Addressing the perfect storm: Biomarkers in obesity and pathophysiology of cardiometabolic risk. *Clinical Chemistry*, *64*(1), 142–153. https://doi.org/10.1373/clinchem.2017.275172

Allin, K. H., Bojesen, S. E., & Nordestgaard, B. G. (2009). Baseline C-reactive protein is associated with incident cancer and survival in patients with cancer. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 27*(13), 2217–2224. https://doi.org/10.1200/JCO.2008.19.8440

Altman, N., & Krzywinski, M. (2015). Points of Significance: Association, correlation and causation. *Nature Methods, 12*, 899–900. https://doi.org/10.1038/nmeth.3587

American Cancer Society. (2015) *Health risks of secondhand smoke: What is secondhand smoke?* Retrieved August 16, 2018, from https://www.cancer.org/cancer/cancer-causes/tobacco-and-cancer/secondhand-smoke.html

American Cancer Society. (2016a). *Small cell lung cancer survival rates, by stage*. Retrieved August 27, 2018, from https://www.cancer.org/cancer/small-cell-lung-cancer/detection-diagnosis-staging/survival-rates.html

American Cancer Society. (2016b). *What is non-small cell lung cancer?* Retrieved August 7, 2018, from https://www.cancer.org/cancer/non-small-cell-lung-cancer/about/what-is-non-small-cell-lung-cancer.html

American Cancer Society. (2017). *Non-small cell lung cancer survival rates, by stage.* Retrieved August 27, 2018, from https://www.cancer.org/cancer/non-small-cell-lung-cancer/detection-diagnosis-staging/survival-rates.html

American Cancer Society. (2019). *Key statistics for small cell lung* cancer. Retrieved June 1, 2019, from https://www.cancer.org/cancer/small-cell-lung-cancer/about/key-statistics.html

American Joint Committee on Cancer. (2018). *What is cancer staging?* Retrieved August 10, 2018, from https://cancerstaging.org/references-tools/pages/What-is-Cancer-Staging.aspx

Argirion, I., Weinstein, S. J., Männistö, S., Albanes, D., & Mondul, A. M. (2017). Serum insulin, glucose, indices of insulin resistance, and risk of lung cancer. *Cancer Epidemiology and Prevention Biomarkers, 26*(10), 1519–1524. https://doi.org/10.1158/1055-9965.EPI-17-0293

Awad, M., & Khanna, R. (2015). Machine learning. In M. Awad & R. Khanna (Eds.), *Efficient learning machines: Theories, concepts, and applications for engineers and system designers* (pp. 1–18). Berkeley, CA: Apress. https://doi.org/10.1007/978-1-4302-5990-9_1

Bae, J.-M., & Kim, E. H. (2015). Hormonal replacement therapy and the risk of lung cancer in women: An adaptive meta-analysis of cohort studies. *Journal of Preventive Medicine and Public Health, 48*(6), 280–286. https://doi.org/10.3961/jpmph.15.054

Bassuk, S. S., Rifai, N., & Ridker, P. M. (2004). High-sensitivity C-reactive protein: Clinical importance. *Current Problems in Cardiology*, *29*(8), 439–493.

Bayliss, R., Choi, J., Fennell, D. A., Fry, A. M., & Richards, M. W. (2016). Molecular mechanisms that underpin EML4-ALK driven cancers and their response to targeted drugs. *Cellular and Molecular Life Sciences, 73*, 1209–1224. https://doi.org/10.1007/s00018-015-2117-6

Beckles, M. A., Spiro, S. G., Colice, G. L., & Rudd, R. M. (2003). Initial evaluation of the patient with lung cancer: Symptoms, signs, laboratory tests, and paraneoplastic syndromes. *Chest, 123*(1 Suppl), 97S-104S.

Biesbroek, S., A, V. D., L, D., Brosens, M. C., Beulens, J. W., Verschuren, W. M., … Boer, J. M. (2015). Identifying cardiovascular risk factor–related dietary patterns with reduced rank regression and random forest in the EPIC-NL cohort. *The American Journal of Clinical Nutrition, 102*(1), 146–154. https://doi.org/10.3945/ajcn.114.092288

Booth, C. M., Li, G., Zhang-Salomons, J., & Mackillop, W. J. (2010). The impact of socioeconomic status on stage of cancer at diagnosis and survival: A population-based study in Ontario, Canada. *Cancer, 116*(17), 4160–4167. https://doi.org/10.1002/cncr.25427

Borch, K. B., Weiderpass, E., Braaten, T., Hansen, M. S., & Licaj, I. (2018). Risk of lung cancer and physical activity by smoking status and body mass index, the Norwegian Women and Cancer Study. *European Journal of Epidemiology*. https://doi.org/10.1007/s10654-018-0446-0

Brady, W. E., Mares-Perlman, J. A., Bowen, P., & Stacewicz-Sapuntzakis, M. (1996). Human serum carotenoid concentrations are related to physiologic and lifestyle factors. *The Journal of Nutrition, 126*(1), 129–137. https://doi.org/10.1093/jn/126.1.129

Breiman, L. (2001). Random forests. *Machine learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brenner, D. R., Boffetta, P., Duell, E. J., Bickeböller, H., Rosenberger, A., McCormack, V., … Hung, R. J. (2012). Previous lung diseases and lung cancer risk: A pooled analysis from the international lung cancer consortium. *American Journal of Epidemiology, 176*(7), 573–585. https://doi.org/10.1093/aje/kws151

Brenner, D. R., Hung, R. J., Tsao, M.-S., Shepherd, F. A., Johnston, M. R., Narod, S., … McLaughlin, J. R. (2010). Lung cancer risk in never-smokers: a population-based case-control study of epidemiologic risk factors. *BMC Cancer, 10*(285). https://doi.org/10.1186/1471-2407-10-285

Brenner, D. R., McLaughlin, J. R., & Hung, R. J. (2011). Previous lung diseases and lung cancer risk: A systematic review and meta-analysis. *PLoS ONE, 6*(3). https://doi.org/10.1371/journal.pone.0017479

Brenner, D. R., Yannitsos, D. H., Farris, M. S., Johansson, M., & Friedenreich, C. M. (2016). Leisure-time physical activity and lung cancer risk: A systematic review and meta-analysis. *Lung Cancer, 95*, 17–27. https://doi.org/10.1016/j.lungcan.2016.01.021

Psaty, B. M., Dekkers, O. M., & Cooper, R. S. (2018). Comparison of 2 treatment models: Precision medicine and preventive medicine. *JAMA, 320*(8), 751–752. https://doi.org/10.1001/jama.2018.8377

Bursac, Z., Gauss, C. H., Williams, D. K., & Hosmer, D. W. (2008). Purposeful selection of variables in logistic regression. *Source Code for Biology and Medicine, 3*(1), 17. https://doi.org/10.1186/1751-0473-3-17

Canadian Cancer Statistics Advisory Committee. (2018). *Diagnosis of lung cancer.* Retrieved August 18, 2018, from http://www.cancer.ca/en/cancer-information/cancer-type/lung/diagnosis/?region=on

Canadian Cancer Statistics Advisory Committee. (2019). *Survival statistics for non–small cell lung cancer*. Retrieved August 24, 2019, from http://www.cancer.ca website: https://www.cancer.ca:443/en/cancer-information/cancer-type/lung/prognosis-and-survival/non-small-cell-lung-cancer-survival-statistics/?region=ab

Canadian Cancer Statistics Advisory Committee. (2017). *Canadian Cancer Statistics 2017.* Toronto, ON: Canadian Cancer Society; 2018. Available at: cancer.ca/Canadian-Cancer-Statistics-2017-EN (accessed [2018-08-26]).

Canadian Cancer Statistics Advisory Committee. (2018). *Canadian Cancer Statistics 2018.* Toronto, ON: Canadian Cancer Society; 2018. Available at: cancer.ca/Canadian-Cancer-Statistics-2018-EN (accessed [2018-08-27]).

Centers for Disease Control and Prevention. (2010). *Racial/ethnic disparities and geographic differences in lung cancer incidence --- 38 states and the district of columbia, 1998--2006.* Retrieved August 14, 2018, from https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5944a2.htm

Centers for Disease Control and Prevention. (2018). *Pneumonia: Causes of pneumonia.* Retrieved December 5, 2018, from https://www.cdc.gov/pneumonia/causes.html

Chaturvedi, A. K., Caporaso, N. E., Katki, H. A., Wong, H.-L., Chatterjee, N., Pine, S. R., … Engels, E. A. (2010). C-reactive protein and risk of lung cancer. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology, 28*(16), 2719–2726. https://doi.org/10.1200/JCO.2009.27.0454

Cheepsattayakorn, A., & Cheepsattayakorn, R. (2014). Parasitic pneumonia and lung involvement. *BioMed Research International, 2014*. https://doi.org/10.1155/2014/874021

Choi, J., Joseph, L., & Pilote, L. (2013). Obesity and C-reactive protein in various populations: A systematic review and meta-analysis. *Obesity Reviews, 14*(3), 232–244. https://doi.org/10.1111/obr.12003

Chyou, P. H., Nomura, A. M., & Stemmermann, G. N. (1992). A prospective study of the attributable risk of cancer due to cigarette smoking. *American Journal of Public Health, 82*(1), 37–40.

Cibickova, L., Karasek, D., Langova, K., Vaverkova, H., Orsag, J., Lukes, J., & Novotny, D. (2015). Correlation of lipid parameters and markers of insulin resistance: Does smoking make a difference? *Atherosclerosis, 241*(1), e136–e137. https://doi.org/10.1016/j.atherosclerosis.2015.04.473

Clark, D. O., Unroe, K. T., Xu, H., Keith, N. R., Callahan, C. M., & Tu, W. (2016). Sex and race differences in the relationship between obesity and C-reactive protein. *Ethnicity & Disease, 26*(2), 197–204. https://doi.org/10.18865/ed.26.2.197

Collins, L. G., Haines, C., Perkel, R., & Enck, R. E. (2007). Lung cancer: Diagnosis and management. *American Family Physician, 75*(1), 56–63.

Coronel, J., Pinos, I., & Amengual, J. (2019). β-carotene in obesity research: Technical considerations and current status of the field. *Nutrients, 11*(4), 842. https://doi.org/10.3390/nu11040842

Corraini, P., Olsen, M., Pedersen, L., Dekkers, O. M., & Vandenbroucke, J. P. (2017). Effect modification, interaction and mediation: An overview of theoretical insights for clinical investigators. *Clinical Epidemiology, 9*, 331–338. https://doi.org/10.2147/CLEP.S129728

Dalton, S. O., Frederiksen, B. L., Jacobsen, E., Steding-Jessen, M., Østerlind, K., Schüz, J., … Johansen, C. (2011). Socioeconomic position, stage of lung cancer and time between referral and diagnosis in Denmark, 2001–2008. *British Journal of Cancer, 105*(7), 1042–1048. https://doi.org/10.1038/bjc.2011.342

de Groot, P. M., Wu, C. C., Carter, B. W., & Munden, R. F. (2018). The epidemiology of lung cancer. *Translational Lung Cancer Research, 7*(3), 220–233. https://doi.org/10.21037/tlcr.2018.05.06

Deedwania, P. C., & Gupta, R. (2006). Management issues in the metabolic syndrome. *The Journal of the Association of Physicians of India, 54*, 797–810.

Dela Cruz, C. S., Tanoue, L. T., & Matthay, R. A. (2011). Lung Cancer: Epidemiology, Etiology, and Prevention. *Clinics in Chest Medicine, 32*(4), 605–644. https://doi.org/10.1016/j.ccm.2011.09.001

Denholm, R., Schüz, J., Straif, K., Stücker, I., Jöckel, K.-H., Brenner, D. R., … C Olsson, A. (2014). Is previous respiratory disease a risk factor for lung cancer? *American Journal of Respiratory and Critical Care Medicine, 190*(5), 549–559. https://doi.org/10.1164/rccm.201402-0338OC

Detterbeck, F. C., Boffa, D. J., Kim, A. W., & Tanoue, L. T. (2017). The Eighth Edition Lung Cancer Stage Classification. *Chest, 151*(1), 193–203. https://doi.org/10.1016/j.chest.2016.10.010

Doll, R., & Hill, A. B. (1950). Smoking and Carcinoma of the Lung. *British Medical Journal, 2*(4682), 739–748.

Dong, Y., & Peng, C.-Y. J. (2013). Principled missing data methods for researchers. *SpringerPlus, 2*. https://doi.org/10.1186/2193-1801-2-222

Dowling, R. J. O., Niraula, S., Stambolic, V., & Goodwin, P. J. (2012). Metformin in cancer: Translational challenges. *Journal of Molecular Endocrinology, 48*(3), R31–R43. https://doi.org/10.1530/JME-12-0007

Duan, P., Hu, C., Quan, C., Yi, X., Zhou, W., Yuan, M., … Yang, K. (2015). Body mass index and risk of lung cancer: Systematic review and dose-response meta-analysis. *Scientific Reports, 5*. https://doi.org/10.1038/srep16938

Durham, A. L., & Adcock, I. M. (2015). The relationship between COPD and lung cancer. *Lung Cancer, 90*(2), 121–127. https://doi.org/10.1016/j.lungcan.2015.08.017

Eckel, R. H., Grundy, S. M., & Zimmet, P. Z. (2005). The metabolic syndrome. *The Lancet, 365*(9468), 1415–1428. https://doi.org/10.1016/S0140-6736(05)66378-7

Edge, S. B. (2017). *AJCC cancer staging manual.* New York: Springer, c2017. Retrieved from

https://proxy.library.brocku.ca/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cat00778a&AN=bu.b2897605&site=eds-live&scope=site

Egom, E. E., Pharithi, R. B., Shiwani, H. A., Khan, B., Kruzliak, P., El-Hiani, Y., & Maher, V. (2018). Time to redefine body mass index categories in chronic diseases? Spotlight on obesity paradox. *International Journal of Food Sciences and Nutrition, 69*(5), 513–523. https://doi.org/10.1080/09637486.2017.1389859

El-Zein, M., Parent, M., Nicolau, B., Koushik, A., Siemiatycki, J., & Rousseau, M. (2013). Body mass index, lifetime smoking intensity and lung cancer risk. *International Journal of Cancer, 133*(7), 1721–1731. https://doi.org/10.1002/ijc.28185

Epstein, F. H., Gabay, C., & Kushner, I. (1999). Acute-phase proteins and other systemic responses to inflammation. *The New England Journal of Medicine, 340*(6), 448–454. https://doi: 10.1056/NEJM199902113400607

Field, A. P. (2009). *Discovering statistics using SPSS (and sex, drugs and rock "n" roll)*. Los Angeles : SAGE Publications, 2009.

Fung, T. T., Rimm, E. B., Spiegelman, D., Rifai, N., Tofler, G. H., Willett, W. C., & Hu, F. B. (2001). Association between dietary patterns and plasma biomarkers of obesity and cardiovascular disease risk. *The American Journal of Clinical Nutrition, 73*(1), 61–67. https://doi.org/10.1093/ajcn/73.1.61

Gallicchio, L., Boyd, K., Matanoski, G., Tao, X. (Grant), Chen, L., Lam, T. K., … Alberg, A. J. (2008). Carotenoids and the risk of developing lung cancer: A systematic review. *The American Journal of Clinical Nutrition, 88*(2), 372–383. https://doi.org/10.1093/ajcn/88.2.372

Gohagan, J. K., Prorok, P. C., Hayes, R. B., & Kramer, B.-S. (2000). The prostate, lung, colorectal and ovarian (PLCO) cancer screening trial of the national cancer institute: History, organization, and status. *Controlled Clinical Trials, 21*(6), 251S-272S. https://doi.org/10.1016/S0197-

2456(00)00097-0

Goodman, G. E., Thornquist, M. D., Balmes, J., Cullen, M. R., Meyskens, F. L., Omenn, G. S., … Williams, J. H. (2004). The beta-carotene and retinol efficacy trial: Incidence of lung cancer and cardiovascular disease mortality during 6-year follow-up after stopping beta-carotene and retinol supplements. *Journal of the National Cancer Institute, 96*(23), 1743–1750. https://doi.org/10.1093/jnci/djh320

Goodwin, P. J., Tammemagi, M., Stambolic, V., Dowling, R. J., Williams, C., Moore, M., … Shepherd, F. A. (2015). Modifiable metabolic markers C-peptide (C-PEP), highly sensitive C-reactive protein (hsCRP), leptin (LEP)] and lung cancer (LC) risk: A matched case-control study nested in the prostate, lung, colorectal and ovarian (PLCO) cancer screening study. *Journal of Clinical Oncology, 33*(15_suppl), 1520–1520. https://doi.org/10.1200/jco.2015.33.15_suppl.1520 An Abstract, presented in Annual Meeting of the American-Society-of-Clinical-Oncology (ASCO) / Clinical Science Symposium on Predicting and Improving Adverse Outcomes in Older Adults with Cancer; May 29-June 02, 2015; Chicago, Illinois (IL)

Goodwin, P., & Stambolic, V. (2015). Impact of the Obesity Epidemic on Cancer. *Annual Review of Medicine, 66*(1), 281–296. https://doi.org/10.1146/annurev-med-051613-012328

Government of Canada. (2004) *Canadian guidelines for body weight classification in adults: Body mass index (BMI) nomogram.* Retrieved from https://www.canada.ca/en/health-canada/services/food-nutrition/healthy-eating/healthy-weights/canadian-guidelines-body-weight-classification-adults/body-mass-index-nomogram.html

Government of Canada. (2017) *Lung Cancer: What is lung cancer?* Retrieved April 10, 2018, from https://www.canada.ca/en/public-health/services/chronic-diseases/cancer/lung-cancer.html

Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science, 8*(3), 206–213.

https://doi.org/10.1007/s11121-007-0070-9

Greenland, S. (1989). Modeling and variable selection in epidemiologic analysis. *American Journal of Public Health, 79*(3), 340–349.

Greenland, S., Schwartzbaum, J. A., & Finkle, W. D. (2000). Problems due to small samples and sparse data in conditional logistic regression analysis. *American Journal of Epidemiology, 151*(5), 531–539. https://doi.org/10.1093/oxfordjournals.aje.a010240

Greiser, C. M., Greiser, E. M., & Dören, M. (2010). Menopausal hormone therapy and risk of lung cancer—Systematic review and meta-analysis. *Maturitas, 65*(3), 198–204. https://doi.org/10.1016/j.maturitas.2009.11.027

Grund, B., & Sabin, C. (2010). Analysis of biomarker data: Logs, odds ratios and ROC curves. *Current pinion in HIV and AIDS, 5*(6), 473–479. https://doi.org/10.1097/COH.0b013e32833ed742

Hales, C. M., Carroll, M. D., Fryar, C. D., & Ogden, C. L. (2017). Prevalence of obesity among adults and youth: United States, 2015–2016. *Centers for Disease Control and Prevention, National Center for Health Statistics*, (288), 8.

Hamilton, W., Peters, T., Round, A., & Sharp, D. (2005). What are the clinical features of lung cancer before the diagnosis is made? A population based case-control study. *Thorax, 60*(12), 1059–1065. https://doi.org/10.1136/thx.2005.045880

Hayes, R. B., Reding, D., Kopp, W., Subar, A. F., Bhat, N., Rothman, N., … Gohagan, J. K. (2000). Etiologic and early marker studies in the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Controlled Clinical Trials, 21*(6), 349S-355S.

Hebden, L., O'Leary, F., Rangan, A., Lie, E. S., Hirani, V., & Allman-Farinelli, M. (2017). Fruit consumption and adiposity status in adults: A systematic review of current evidence. *Critical Reviews in Food Science and Nutrition, 57*(12), 2526–2540. https://doi.org/10.1080/10408398.2015.1012290

Helmersson, J., Larsson, A., Vessby, B., & Basu, S. (2005). Active smoking and a history of smoking are

associated with enhanced prostaglandin F2α, interleukin-6 and F2-isoprostane formation in elderly men. *Atherosclerosis, 181*(1), 201–207. https://doi.org/10.1016/j.atherosclerosis.2004.11.026

Herbst, R. S., Heymach, J. V., & Lippman, S. M. (2008). Lung Cancer. *New England Journal of Medicine, 359*(13), 1367–1380. https://doi.org/10.1056/NEJMra0802714

Herth, F. J. F., Eberhardt, R., Vilmann, P., Krasnik, M., & Ernst, A. (2006). Real-time endobronchial ultrasound guided transbronchial needle aspiration for sampling mediastinal lymph nodes. *Thorax, 61*(9), 795–798. https://doi.org/10.1136/thx.2005.047829

Hoffman, P. C., Mauer, A. M., & Vokes, E. E. (2000). Lung cancer. *The Lancet, 355*(9202), 479–485. https://doi.org/10.1016/S0140-6736(00)82038-3

Hopkins, B. D., Goncalves, M. D., & Cantley, L. C. (2016). Obesity and cancer mechanisms: Cancer metabolism. *Journal of Clinical Oncology, 34*(35), 4277–4283. https://doi.org/10.1200/JCO.2016.67.9712

Hori, M., Tanaka, H., Wakai, K., Sasazuki, S., & Katanoda, K. (2016). Secondhand smoke exposure and risk of lung cancer in Japan: A systematic review and meta-analysis of epidemiologic studies. *Japanese Journal of Clinical Oncology, 46*(10), 942–951. https://doi.org/10.1093/jjco/hyw091

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. New York: Wiley, c2000.

Howington, J. A., Blum, M. G., Chang, A. C., Balekian, A. A., & Murthy, S. C. (2013). Treatment of stage I and II non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest, 143*(5), e278S-e313S. https://doi.org/10.1378/chest.12-2359

Hsu, C.-N., Chang, C.-H., Lin, Y.-S., Lin, J.-W., & Caffrey, J. L. (2013). Association of serum C-peptide concentrations with cancer mortality risk in pre-diabetes or undiagnosed diabetes. *PloS One, 8*(2), e55625. https://doi.org/10.1371/journal.pone.0055625

Hu, F. B., Rimm, E., Smith-Warner, S. A., Feskanich, D., Stampfer, M. J., Ascherio, A., … Willett, W. C. (1999). Reproducibility and validity of dietary patterns assessed with a food-frequency questionnaire. *The American Journal of Clinical Nutrition, 69*(2), 243–249. https://doi.org/10.1093/ajcn/69.2.243

International Agency for Research on Cancer [IARC]. (2018). *Cancer Today*. Lyon, France: World Health Organization. Retrieved from https://gco.iarc.fr/today/home

Jamalan, M., Rezazadeh, M., Zeinali, M., & Ghaffari, M. A. (2015). Effect of ascorbic acid and alpha-tocopherol supplementations on serum leptin, tumor necrosis factor alpha, and serum amyloid A levels in individuals with type 2 diabetes mellitus. *Avicenna Journal of Phytomedicine, 5*(6), 531–539.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: with applications in R (Vol. 103)*. New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-7138-7

Jemal, A., Center, M. M., & Ward, E. (2009). The convergence of lung cancer rates between blacks and whites under the age of 40, United States. *Cancer Epidemiology and Prevention Biomarkers, 18*(12), 3349–3352. https://doi.org/10.1158/1055-9965.EPI-09-0740

Jett, J. R., Schild, S. E., Kesler, K. A., & Kalemkerian, G. P. (2013). Treatment of small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest, 143*(5), e400S-e419S. https://doi.org/10.1378/chest.12-2363

Jones, A. G., & Hattersley, A. T. (2013). The clinical utility of C-peptide measurement in the care of patients with diabetes. *Diabetic Medicine, 30*(7), 803–817. https://doi.org/10.1111/dme.12159

Julia, C., Meunier, N., Touvier, M., Ahluwalia, N., Sapin, V., Papet, I., … Kesse-Guyot, E. (2013). Dietary patterns and risk of elevated C-reactive protein concentrations 12 years later. *British Journal of*

Nutrition, 110(04), 747–754. https://doi.org/10.1017/S0007114512005636

Kabat, G. C., Kim, M. Y., Hollenbeck, A. R., & Rohan, T. E. (2014). Attained height, sex, and risk of cancer at different anatomic sites in the NIH-AARP diet and health study. *Cancer Causes & Control: CCC, 25*(12), 1697–1706. https://doi.org/10.1007/s10552-014-0476-1

Kamath, D. Y., Xavier, D., Sigamani, A., & Pais, P. (2015). High sensitivity C-reactive protein (hsCRP) & cardiovascular disease: An Indian perspective. *The Indian Journal of Medical Research, 142*(3), 261–268. https://doi.org/10.4103/0971-5916.166582

Kay, F. U., Kandathil, A., Batra, K., Saboo, S. S., Abbara, S., & Rajiah, P. (2017). Revisions to the tumor, node, metastasis staging of lung cancer (8th edition): Rationale, radiologic findings and clinical implications. *World Journal of Radiology, 9*(6), 269–279. https://doi.org/10.4329/wjr.v9.i6.269

Kazmi, A., Sattar, A., Hashim, R., Khan, S. P., Younus, M., & Khan, F. A. (2013). Serum Leptin values in the healthy obese and non-obese subjects of Rawalpindi. *Journal of Pakistan Medical Association, 63*(2), 245–248.

Khuder, S. A., Dayal, H. H., Mutgi, A. B., Willey, J. C., & Dayal, G. (1998). Effect of cigarette smoking on major histological types of lung cancer in men. *Lung Cancer, 22*(1), 15–21. https://doi.org/10.1016/S0169-5002(98)00068-3

Kligerman, S., & White, C. (2011). Epidemiology of lung cancer in women: Risk factors, survival, and screening. *American Journal of Roentgenology, 196*(2), 287–295. https://doi.org/10.2214/AJR.10.5412

Klok, M. D., Jakobsdottir, S., & Drent, M. L. (2007). The role of leptin and ghrelin in the regulation of food intake and body weight in humans: a review. *Obesity Reviews, 8*(1), 21–34. https://doi.org/10.1111/j.1467-789X.2006.00270.x

Knol, M. J., Egger, M., Scott, P., Geerlings, M. I., & Vandenbroucke, J. P. (2009). When one depends on the other: Reporting of interaction in case-control and cohort studies. *Epidemiology, 20*(2),

161–166. https://doi.org/10.1097/EDE.0b013e31818f6651

Ko, B.J., Park, K. H., Shin, S., Zaichenko, L., Davis, C. R., Crowell, J. A., … Mantzoros, C. S. (2016). Diet quality and diet patterns in relation to circulating cardiometabolic biomarkers. *Clinical Nutrition, 35*(2), 484–490. https://doi.org/10.1016/j.clnu.2015.03.022

Kollarova, H., Machova, L., Horakova, D., Cizek, L., Janoutova, G., & Janout, V. (2008). Is obesity a preventive factor for lung cancer? *Neoplasma, 55*(1), 71—73

Kovalchik, S. A., Tammemagi, M., Berg, C. D., Caporaso, N. E., Riley, T. L., Korch, M., … Katki, H. A. (2013). Targeting of low-dose CT screening according to the risk of lung-cancer death. *The New England Journal of Medicine*, *369*(3), 245–254. https://doi.org/10.1056/NEJMoa1301851

Lam, T. K., Gallicchio, L., Lindsley, K., Shiels, M., Hammond, E., Tao, X. G., … Alberg, A. J. (2009). Cruciferous vegetable consumption and lung cancer risk: A systematic review. *Cancer Epidemiology, Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology, 18*(1), 184–195. https://doi.org/10.1158/1055-9965.EPI-08-0710

Lam, T. K., Ruczinski, I., Helzlsouer, K. J., Shugart, Y. Y., Caulfield, L. E., & Alberg, A. J. (2010). Cruciferous Vegetable Intake and Lung Cancer Risk: A Nested Case-Control Study Matched on Cigarette Smoking. *Cancer Epidemiology and Prevention Biomarkers, 19*(10), 2534–2540. https://doi.org/10.1158/1055-9965.EPI-10-0475

Lamprecht, B., Porsch, P., Pirich, C., & Studnicka, M. (2009). Electromagnetic Navigation Bronchoscopy in Combination with PET-CT and Rapid On-site Cytopathologic Examination for Diagnosis of Peripheral Lung Lesions. *Lung, 187*(1), 55. https://doi.org/10.1007/s00408-008-9120-8

Landi, M. T., Chatterjee, N., Yu, K., Goldin, L. R., Goldstein, A. M., Rotunno, M., … Caporaso, N. E. (2009). A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *American Journal of Human Genetics, 85*(5), 679–691.

https://doi.org/10.1016/j.ajhg.2009.09.012

Latimer, K., & Mott, T. (2015). Lung cancer: Diagnosis, treatment principles, and screening. *American Family Physician, 91*(4), 250–256.

Lee, P. N., Forey, B. A., & Coombs, K. J. (2012). Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. *BMC Cancer, 12*(385). https://doi.org/10.1186/1471-2407-12-385

Lewis, D. R., Check, D. P., Caporaso, N. E., Travis, W. D., & Devesa, S. S. (2014). US lung cancer trends by histologic type. *Cancer, 120*(18), 2883–2892. https://doi.org/10.1002/cncr.28749

Li, M., Fan, Y., Zhang, X., Hou, W., & Tang, Z. (2014). Fruit and vegetable intake and risk of type 2 diabetes mellitus: Meta-analysis of prospective cohort studies. *BMJ Open, 4*(11), e005497. https://doi.org/10.1136/bmjopen-2014-005497

Lin, J., Cook, N. R., Albert, C., Zaharris, E., Gaziano, J. M., Van Denburgh, M., … Manson, J. E. (2009). Vitamins C and E and beta carotene supplementation and cancer risk: A randomized controlled trial. *JNCI: Journal of the National Cancer Institute, 101*(1), 14–23. https://doi.org/10.1093/jnci/djn438

Lin, J., Gill, A., Zahm, S. H., Carter, C. A., Shriver, C. D., Nations, J. A., … Zhu, K. (2017). Metformin use and survival after non-small cell lung cancer: A cohort study in the US Military health system. *International Journal of Cancer, 141*(2), 254–263. https://doi.org/10.1002/ijc.30724

Lips, E. H., Gaborieau, V., McKay, J. D., Chabrier, A., Hung, R. J., Boffetta, P., … Brennan, P. (2010). Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17 000 individuals. *International Journal of Epidemiology, 39*(2), 563–577. https://doi.org/10.1093/ije/dyp288

Lissowska, J., Foretova, L., Dąbek, J., Zaridze, D., Szeszenia-Dabrowska, N., Rudnai, P., … Boffetta, P. (2010). Family history and lung cancer risk: International multicentre case–control study in

Eastern and Central Europe and meta-analyses. *Cancer Causes & Control, 21*(7), 1091–1104. https://doi.org/10.1007/s10552-010-9537-2

Little, T. D., Jorgensen, T. D., Lang, K. M., & Moore, E. W. G. (2014). On the joys of missing data. *Journal of Pediatric Psychology, 39*(2), 151–162. https://doi.org/10.1093/jpepsy/jst048

Lubin, J. H., & Caporaso, N. E. (2006). Cigarette smoking and lung cancer: Modeling total exposure and intensity. *Cancer Epidemiology and Prevention Biomarkers, 15*(3), 517–523. https://doi.org/10.1158/1055-9965.EPI-05-0863

Malley, J. D., Malley, K. G., & Pajevic, S. (2011). *Practical guides to biostatistics and epidemiology: Statistical learning for biomedical data (1st ed).* Cambridge: Cambridge University Press.

Mansournia, M. A., Jewell, N. P., & Greenland, S. (2018). Case—control matching: Effects, misconceptions, and recommendations. *European Journal of Epidemiology, (1)*5. https://doi.org/10.1007/s10654-017-0325-0

Mariotto, A. B., Noone, A.-M., Howlader, N., Cho, H., Keel, G. E., Garshell, J., … Schwartz, L. M. (2014). Cancer survival: An overview of measures, uses, and interpretation. *Journal of the National Cancer Institute. Monographs, 2014*(49), 145–186. https://doi.org/10.1093/jncimonographs/lgu024

Mazzone, P. J., Rai, H., Beukemann, M., Xu, M., Jain, A., & Sasidhar, M. (2012). The effect of metformin and thiazolidinedione use on lung cancer in diabetics. *BMC Cancer, 12*(1). https://doi.org/10.1186/1471-2407-12-410

McAuley, P. A., & Blair, S. N. (2011). Obesity paradoxes. *Journal of Sports Sciences*, 29(8), 773–782. https://doi.org/10.1080/02640414.2011.553965

McCarthy, W. J., Meza, R., Jeon, J., & Moolgavkar, S. (2012). Lung cancer in never smokers epidemiology and risk prediction models. *Risk Analysis, 32*(Suppl 1), S69–S84. https://doi.org/10.1111/j.1539-6924.2012.01768.x

Medina-Remón, A., Kirwan, R., Lamuela-Raventós, R. M., & Estruch, R. (2018). Dietary patterns and the risk of obesity, type 2 diabetes mellitus, cardiovascular diseases, asthma, and neurodegenerative diseases. *Critical Reviews in Food Science and Nutrition, 58*(2), 262–296. https://doi.org/10.1080/10408398.2016.1158690

Mirsadraee, S., Oswal, D., Alizadeh, Y., Caulo, A., & van Beek, E. J. (2012). The 7th lung cancer TNM classification and staging system: Review of the changes and implications. *World Journal of Radiology, 4*(4), 128–134. https://doi.org/10.4329/wjr.v4.i4.128

Mitra, D., Shaw, A., & Tjepkema, M. (2015). Social determinants of lung cancer incidence in Canada: A 13-year prospective study. *Health Reports, 26*(6), 11.

Molina, J. R., Yang, P., Cassivi, S. D., Schild, S. E., & Adjei, A. A. (2008). Non–small cell lung cancer: Epidemiology, risk factors, treatment, and survivorship. Mayo Clinic Proceedings. *Mayo Clinic, 83*(5), 584–594.

Morgillo, F., Sasso, F. C., Corte, C. M. D., Festino, L., Manzo, A., Martinelli, E., … Ciardiello, F. (2013). Metformin in lung cancer: Rationale for a combination therapy. *Expert Opinion on Investigational Drugs, 22*(11), 1401–1409. https://doi.org/10.1517/13543784.2013.828691

Müller, F. H. (1940). Tabakmißbrauch und Lungencarcinom. *Zeitschrift für Krebsforschung, 49*(1), 57–85. https://doi.org/10.1007/BF01633114

Münzberg, H., & Morrison, C. D. (2015). Structure, production and signaling of leptin. *Metabolism: Clinical and Experimental, 64*(1), 13–23. https://doi.org/10.1016/j.metabol.2014.09.010

National Cancer Institute. (2013). *Diagnosis and staging: Tumor grade.* Retrieved August 11, 2018, from https://www.cancer.gov/about-cancer/diagnosis-staging/prognosis/tumor-grade-fact-sheet

National Cancer Institute. (2015a) *Diagnosis and staging: Cancer staging.* Retrieved August 8, 2018, from https://www.cancer.gov/about-cancer/diagnosis-staging/staging

National Cancer Institute. (2015b). Understanding Cancer: *What is cancer?* Retrieved August 19, 2018,

from https://www.cancer.gov/about-cancer/understanding/what-is-cancer

National Center for Health Statistics. (2017). Leading causes of death in the United States. Retrieved August 27, 2019, from https://www.cdc.gov/nchs/fastats/leading-causes-of-death.htm

Nettleton, J. A., Steffen, L. M., Mayer-Davis, E. J., Jenny, N. S., Jiang, R., Herrington, D. M., & Jacobs, D. R. (2006). Dietary patterns are associated with biochemical markers of inflammation and endothelial activation in the Multi-Ethnic Study of Atherosclerosis (MESA). *The American Journal of Clinical Nutrition, 83*(6), 1369–1379.

Nimptsch, K., & Pischon, T. (2016). Obesity biomarkers, metabolism and risk of cancer: An epidemiological perspective. Recent results in cancer research. *Fortschritte Der Krebsforschung.* Progres Dans Les Recherches Sur Le Cancer, 208, 199–217. https://doi.org/10.1007/978-3-319-42542-9_11

North, C. M., & Christiani, D. C. (2013). Women and lung cancer: What's new? *Seminars in Thoracic and Cardiovascular Surgery, 25*(2). https://doi.org/10.1053/j.semtcvs.2013.05.002

Ntikoudi, E., Kiagia, M., Boura, P., & Syrigos, K. N. (2014). Hormones of adipose tissue and their biologic role in lung cancer. *Cancer Treatment Reviews, 40*(1), 22–30. https://doi.org/10.1016/j.ctrv.2013.06.005

Nuttall, F. Q. (2015). Body mass index. *Nutrition Today, 50*(3), 117–128. https://doi.org/10.1097/NT.0000000000000092

Oken, M. M., Hocking, W. G., Kvale, P. A., Andriole, G. L., Buys, S. S., Church, T. R., … Team, for the P. P. (2011). Screening by chest radiograph and lung cancer mortality: The prostate, lung, colorectal, and ovarian (PLCO) randomized trial. *JAMA, 306*(17), 1865–1873. https://doi.org/10.1001/jama.2011.1591

Oken, M. M., Marcus, P. M., Hu, P., Beck, T. M., Hocking, W., Kvale, P. A., … Gohagan, J. K. (2005). Baseline chest radiograph for lung cancer detection in the randomized prostate, lung,

colorectal and ovarian cancer screening trial. *Journal of the National Cancer Institute, 97*(24), 1832–1839. https://doi.org/10.1093/jnci/dji430

O'Neill, S., & O'Driscoll, L. (2015). Metabolic syndrome: A closer look at the growing epidemic and its associated pathologies. *Obesity Reviews, 16*(1), 1–12. https://doi.org/10.1111/obr.12229

Östh, M., Öst, A., Kjolhede, P., & Strålfors, P. (2014). The concentration of β-carotene in human adipocytes, but not the whole-body adipocyte stores, is reduced in obesity. *PLOS ONE, 9*(1), e85610. https://doi.org/10.1371/journal.pone.0085610

Panaretos, D., Koloverou, E., Dimopoulos, A. C., Kouli, G.-M., Vamvakari, M., Tzavelas, G., … Panagiotakos, D. B. (2018). A comparison of statistical and machine-learning techniques in evaluating the association between dietary patterns and 10-year cardiometabolic risk (2002–2012): The ATTICA study. *British Journal of Nutrition, 120*(3), 326–334. https://doi.org/10.1017/S0007114518001150

Patten, S. (2015). *Epidemiology for Canadian students: Principles, methods and critical appraisal*. Brush Education.

Pearce, N. (2016). Analysis of matched case-control studies. *BMJ, 352*, i969. https://doi.org/10.1136/bmj.i969

Pedersen, A. B., Mikkelsen, E. M., Cronin-Fenton, D., Kristensen, N. R., Pham, T. M., Pedersen, L., & Petersen, I. (2017). Missing data and multiple imputation in clinical epidemiological research. *Clinical Epidemiology, 9*, 157–166. https://doi.org/10.2147/CLEP.S129785

Perkins, K. A., & Fonte, C. (2002). Effects of smoking status and smoking cessation on leptin levels. *Nicotine & Tobacco Research, 4*(4), 459–466. https://doi.org/10.1080/1462220021000018434

Pesch, B., Kendzia, B., Gustavsson, P., Jöckel, K.-H., Johnen, G., Pohlabeln, H., … Brüning, T. (2012). Cigarette smoking and lung cancer – relative risk estimates for the major histological types from a pooled analysis of case-control studies. International Journal of Cancer. *Journal International*

Du Cancer, 131(5), 1210–1219. https://doi.org/10.1002/ijc.27339

Pine, S. R., Mechanic, L. E., Enewold, L., Chaturvedi, A. K., Katki, H. A., Zheng, Y.-L., … Harris, C. C. (2011). Increased Levels of Circulating Interleukin 6, Interleukin 8, C-Reactive Protein, and Risk of Lung Cancer. JNCI: Journal of the National Cancer Institute, 103(14), 1112–1122. https://doi.org/10.1093/jnci/djr216

Pinsky, P. F., Miller, A., Kramer, B. S., Church, T., Reding, D., Prorok, P., … Berg, C. D. (2007). Evidence of a healthy volunteer effect in the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial. *American Journal of Epidemiology*, 165(8), 874–881. https://doi.org/10.1093/aje/kwk075

Poirier, A. E., Ruan, Y., Hebert, L. A., Grevers, X., Walter, S. D., Villeneuve, P. J., … Friedenreich, C. M. (2019). Estimates of the current and future burden of cancer attributable to low fruit and vegetable consumption in Canada. *Preventive Medicine*, *122*, 20–30. https://doi.org/10.1016/j.ypmed.2019.03.013

Proctor, R. N. (2012). The history of the discovery of the cigarette–lung cancer link: Evidentiary traditions, corporate denial, global toll. *Tobacco Control, 21*(2), 87–91. https://doi.org/10.1136/tobaccocontrol-2011-050338

Prorok, P. C., Andriole, G. L., Bresalier, R. S., Buys, S. S., Chia, D., David Crawford, E., … Gohagan, J. K. (2000). Design of the prostate, lung, colorectal and ovarian (PLCO) cancer screening trial. *Controlled Clinical Trials, 21*(6), 273S-309S.Pryor, R., & Cabreiro, F. (2015). Repurposing metformin: An old drug with new tricks in its binding pockets. *Biochemical Journal, 471*(Pt 3), 307–322. https://doi.org/10.1042/BJ20150497

Ramnath, N., Dilling, T. J., Harris, L. J., Kim, A. W., Michaud, G. C., Balekian, A. A., … Arenberg, D. A. (2013). Treatment of stage III non-small cell lung cancer: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest, 143*(5), e314S-e340S. https://doi.org/10.1378/chest.12-2360

Reedy, J., Wirfält, E., Flood, A., Mitrou, P. N., Krebs-Smith, S. M., Kipnis, V., … Subar, A. F. (2010). Comparing 3 dietary pattern methods—cluster analysis, factor analysis, and index analysis—with colorectal cancer risk. The NIH–AARP diet and health study. *American Journal of Epidemiology, 171*(4), 479–487. https://doi.org/10.1093/aje/kwp393

Renehan, A. G., Tyson, M., Egger, M., Heller, R. F., & Zwahlen, M. (2008). Body-mass index and incidence of cancer: A systematic review and meta-analysis of prospective observational studies. *The Lancet, 371*(9612), 569–578. https://doi.org/10.1016/S0140-6736(08)60269-X

Ribeiro, R., Araújo, A., Lopes, C., & Medeiros, R. (2007). Immunoinflammatory mechanisms in lung cancer development: Is leptin a mediator? *Journal of Thoracic Oncology, 2*(2), 105–108. https://doi.org/10.1016/S1556-0864(15)30035-6

Richards, T. B., Henley, S. J., Puckett, M. C., Weir, H. K., Huang, B., Tucker, T. C., & Allemani, C. (2017). Lung cancer survival in the United States by race and stage (2001-2009): Findings from the CONCORD-2 study. *Cancer, 123*(S24), 5079–5099. https://doi.org/10.1002/cncr.31029

Rivera, M. P., Mehta, A. C., & Wahidi, M. M. (2013). Establishing the diagnosis of lung cancer: Diagnosis and management of lung cancer, 3rd ed: American college of chest physicians evidence-based clinical practice guidelines. *Chest, 143*(5, Supplement), e142S-e165S. https://doi.org/10.1378/chest.12-2353

Rothman, K.J., Greenland, S, & Lash, T.L. (2008) Design strategies to improve study accuracy. In Rothman, K.J., Greenland, S., & Lash, T.L. (Eds.), *Modern epidemiology* (pp. 168–82). Philadelphia, PA: Lippincott Williams and Wilkins.

Rubin, L. P., Ross, A. C., Stephensen, C. B., Bohn, T., & Tanumihardjo, S. A. (2017). Metabolic effects of inflammation on vitamin A and carotenoids in humans and animal models. *Advances in Nutrition, 8*(2), 197–212. https://doi.org/10.3945/an.116.014167

Ryerson, A. B., Eheman, C. R., Altekruse, S. F., Ward, J. W., Jemal, A., Sherman, R. L., … Kohler, B. A.

(2016). Annual report to the nation on the status of cancer, 1975–2012, featuring the increasing incidence of liver cancer. *Cancer, 122*(9), 1312–1337. https://doi.org/10.1002/cncr.29936

Sagerup, C. M. T., Småstuen, M., Johannesen, T. B., Helland, Å., & Brustugun, O. T. (2011). Sex-specific trends in lung cancer incidence and survival: A population study of 40 118 cases. *Thorax, 66*(4), 301–307. https://doi.org/10.1136/thx.2010.151621

Sanikini, H., Yuan, J.-M., Butler, L. M., Koh, W.-P., Gao, Y.-T., Steffen, A., … Stücker, I. (2018). Body mass index and lung cancer risk: A pooled analysis based on nested case-control studies from four cohort studies. *BMC Cancer, 18*. https://doi.org/10.1186/s12885-018-4124-0

Saracci, R. (1987). The interactions of tobacco smoking and other agents in cancer etiology. *Epidemiologic Reviews, 9*(1), 175–193. https://doi.org/10.1093/oxfordjournals.epirev.a036301

Sauerbrei, W., Meier-Hirmer, C., Benner, A., & Royston, P. (2006). Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. *Computational Statistics & Data Analysis, 50*(12), 3464–3485. https://doi.org/10.1016/j.csda.2005.07.015

Schabath, M. B., Cress, W. D., & Muñoz-Antonia, T. (2016). Racial and ethnic differences in the epidemiology of lung cancer and the lung cancer genome. *Cancer Control: Journal of the Moffitt Cancer Center, 23*(4), 338–346.

Schisterman, E. F., Perkins, N. J., Mumford, S. L., Ahrens, K. A., & Mitchell, E. M. (2017). Collinearity and causal diagrams – a lesson on the importance of model specification. Epidemiology, 28(1): 47–53. doi:10.1097/EDE.0000000000000554.

Schwartz, A. G., & Cote, M. L. (2016). Epidemiology of lung cancer. In A. Ahmad & S. Gadgeel (Eds.), *Lung cancer and personalized medicine: Current knowledge and therapies* (pp. 21–41). Springer, Cham. https://doi.org/10.1007/978-3-319-24223-1_2

Sedgwick, P. (2014). Nested case-control studies: Advantages and disadvantages. *BMJ, 348*, g1532. https://doi.org/10.1136/bmj.g1532

Shiels, M. S., Pfeiffer, R. M., Hildesheim, A., Engels, E. A., Kemp, T. J., Park, J.-H., … Chaturvedi, A. K. (2013). Circulating inflammation markers and prospective risk for lung cancer. *Journal of the National Cancer Institute, 105*(24), 1871–1880. https://doi.org/10.1093/jnci/djt309

Shimazu, T., Inoue, M., Sasazuki, S., Iwasaki, M., Sawada, N., Yamaji, T., & Tsugane, S. (2011). Plasma isoflavones and the risk of lung cancer in women: A nested case–control study in Japan. *Cancer Epidemiology and Prevention Biomarkers, 20*(3), 419–427. https://doi.org/10.1158/1055-9965.EPI-10-1025

Sidorchuk, A., Agardh, E. E., Aremu, O., Hallqvist, J., Allebeck, P., & Moradi, T. (2009). Socioeconomic differences in lung cancer incidence: A systematic review and meta-analysis. *Cancer Causes & Control, 20*(4), 459. https://doi.org/10.1007/s10552-009-9300-8

Siegel, R. L., Miller, K. D., & Jemal, A. (2018). Cancer statistics, 2018. *CA: A Cancer Journal for Clinicians, 68*(1), 7–30. https://doi.org/10.3322/caac.21442Sleire, L., Førde, H. E., Netland, I. A., Leiss, L., Skeie, B. S., & Enger, P. Ø. (2017). Drug repurposing in cancer. *Pharmacological Research, 124*(Complete), 74–91. https://doi.org/10.1016/j.phrs.2017.07.013

Smidowicz, A., & Regula, J. (2015). Effect of nutritional status and dietary patterns on human serum c-reactive protein and interleukin-6 concentrations. *Advances in Nutrition, 6*(6), 738–747. https://doi.org/10.3945/an.115.009415

Soo, R. A., Loh, M., Mok, T. S., Ou, S.-H. I., Cho, B.-C., Yeo, W.-L., … Soong, R. (2011). Ethnic differences in survival outcome in patients with advanced stage non-small cell lung cancer: Results of a meta-analysis of randomized controlled trials. *Journal of Thoracic Oncology, 6*(6), 1030–1038. https://doi.org/10.1097/JTO.0b013e3182199c03

Sookoian, S., & Pirola, C. J. (2011). Metabolic syndrome: From the genetics to the pathophysiology. *Current Hypertension Reports, 13*(2), 149–157. https://doi.org/10.1007/s11906-010-0164-9

Statistics Canada. (2017). Leading causes of death, total population, by age group in Canada. Retrieved Auguest 27, 2019, from https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1310039401

Stark, S. W., RN, APRN, DNSc. (2013). Chronic obstructive pulmonary disease (COPD). Magill's Medical Guide (Online Edition).

Stoltzfus, J. C. (2011). Logistic regression: A brief primer. *Academic Emergency Medicine, 18*(10), 1099–1104. https://doi.org/10.1111/j.1553-2712.2011.01185.x

Tammemägi, M. C., Church, T. R., Hocking, W. G., Silvestri, G. A., Kvale, P. A., Riley, T. L., … Berg, C. D. (2014). Evaluation of the lung cancer risks at which to screen ever- and never-smokers: screening rules applied to the PLCO and NLST cohorts. *PLoS Medicine, 11*(12), e1001764. https://doi.org/10.1371/journal.pmed.1001764

Tammemägi, M. C., Katki, H. A., Hocking, W. G., Church, T. R., Caporaso, N., Kvale, P. A., … Berg, C. D. (2013). Selection criteria for lung-cancer screening. *New England Journal of Medicine, 368*(8), 728–736. https://doi.org/10.1056/NEJMoa1211776

Tammemagi, M. C., Pinsky, P. F., Caporaso, N. E., Kvale, P. A., Hocking, W. G., Church, T. R., … Prorok, P. C. (2011). Lung cancer risk prediction: Prostate, lung, colorectal and ovarian cancer screening trial models and validation. *Journal of the National Cancer Institute, 103*(13), 1058–1068. https://doi.org/10.1093/jnci/djr173

Tammemagi, M. C., Schmidt, H., Martel, S., McWilliams, A., Goffin, J. R., Johnston, M. R., … Lam, S. (2017). Participant selection for lung cancer screening by risk modelling (the Pan-Canadian Early Detection of Lung Cancer [PanCan] study): A single-arm, prospective study. *The Lancet Oncology*, *18*(11), 1523–1531. https://doi.org/10.1016/S1470-2045(17)30597-1

Targher, G., Zenari, L., Faccini, G., Falezza, G., Muggeo, M., & Zoppini, G. (2001). Serum leptin

concentrations in young smokers with type 1 diabetes. *Diabetes Care, 24*(4), 793–794. https://doi.org/10.2337/diacare.24.4.793

Taylor, R., Najafi, F., & Dobson, A. (2007). Meta-analysis of studies of passive smoking and lung cancer: Effects of study type and continent. *International Journal of Epidemiology, 36*(5), 1048–1059. https://doi.org/10.1093/ije/dym158

Temel, J. S., Greer, J. A., Muzikansky, A., Gallagher, E. R., Admane, S., Jackson, V. A., … Lynch, T. J. (2010). Early Palliative Care for Patients with Metastatic Non–Small-Cell Lung Cancer. *New England Journal of Medicine, 363*(8), 733–742. https://doi.org/10.1056/NEJMoa1000678

Thun, M. J., Henley, S. J., Burns, D., Jemal, A., Shanks, T. G., & Calle, E. E. (2006). Lung cancer death rates in lifelong nonsmokers. *Journal of the National Cancer Institute, 98*(10), 691–699. https://doi.org/10.1093/jnci/djj187

Tibuakuu, M., Kamimura, D., Kianoush, S., DeFilippis, A. P., Al Rifai, M., Reynolds, L. M., … Blaha, M. J. (2017). The association between cigarette smoking and inflammation: The genetic epidemiology network of arteriopathy (GENOA) study. *PloS One, 12*(9), e0184914. https://doi.org/10.1371/journal.pone.0184914

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J., & Jemal, A. (2015). Global cancer statistics, 2012. CA: *A Cancer Journal for Clinicians, 65*(2), 87–108. https://doi.org/10.3322/caac.21262

Torre, L. A., Siegel, R. L., & Jemal, A. (2016). Lung cancer statistics. In A. Ahmad & S. Gadgeel (Eds.), *Lung cancer and personalized medicine: Current knowledge and therapies* (pp. 1–19). Location: Springer, Cham. https://doi.org/10.1007/978-3-319-24223-1_1

Tota, J., Ramanakumar, A., & Franco, E. (2014). Lung cancer screening: Review and performance comparison under different risk scenarios. *Lung, 192*(1), 55–63. https://doi.org/10.1007/s00408-013-9517-x

Toyooka, S., Tsuda, T. and Gazdar, A. F. (2003), The *TP53* gene, tobacco exposure, and lung cancer.

Human Mutation, 21, 229-239. doi:10.1002/humu.10177

Travis, W. D. (2011). Pathology of Lung Cancer. *Clinics in Chest Medicine, 32*(4), 669–692. https://doi.org/10.1016/j.ccm.2011.08.005

Trumbo, P., Yates, A. A., Schlicker, S., & Poos, M. (2001). Dietary reference intakes: Vitamin A, vitamin K, arsenic, boron, chromium, copper, iodine, iron, manganese, molybdenum, nickel, silicon, vanadium, and zinc. *Journal of the American Dietetic Association, 101*(3), 294–301. https://doi.org/10.1016/S0002-8223(01)00078-5

Turner, M. C., Chen, Y., Krewski, D., Calle, E. E., & Thun, M. J. (2007). Chronic obstructive pulmonary disease is associated with lung cancer mortality in a prospective study of never smokers. *American Journal of Respiratory and Critical Care Medicine, 176*(3), 285–290. https://doi.org/10.1164/rccm.200612-1792OC

Underwood, J. M., Townsend, J. S., Tai, E., Davis, S. P., Stewart, S. L., White, A., … Fairley, T. L. (2012). Racial and regional disparities in lung cancer incidence. *Cancer, 118*(7), 1910–1918. https://doi.org/10.1002/cncr.26479

United States Census Bureau, Population Division (2019). Annual estimates of the resident population by sex, race, and Hispanic origin for the United States, states, and counties: April 1, 2010 to July 1, 2018

United States Census Bureau, Population Division (2017). Educational attainment of the population 18 years and over, by age, sex, race, and hispanic origin: 2017.

U.S. Department of Agriculture, Agricultural Research Service. What We Eat in America, 2015-2016

U.S. Public Health Service, & National Clearinghouse for Smoking Health. (1972). *The health consequences of smoking: A report of the surgeon general.* Retrieved August 16, 2018, from https://profiles.nlm.nih.gov/NN/B/B/P/M/

Vieira, A. R., Abar, L., Vingeliene, S., Chan, D. S. M., Aune, D., Navarro-Rosenblatt, D., … Norat, T. (2016).

Fruits, vegetables and lung cancer risk: A systematic review and meta-analysis. *Annals of Oncology, 27*(1), 81–96. https://doi.org/10.1093/annonc/mdv381

Virtamo, J., Pietinen, P., Huttunen, J. K., Korhonen, P., Malila, N., Virtanen, M. J., … ATBC Study Group. (2003). Incidence of cancer and mortality following alpha-tocopherol and beta-carotene supplementation: A postintervention follow-up. *JAMA, 290(*4), 476–485. https://doi.org/10.1001/jama.290.4.476

Visser M, Bouter LM, McQuillan GM, Wener MH, Harris TB, Visser, M., … Harris, T. B. (1999). Elevated C-reactive protein levels in overweight and obese adults. *Journal of the American Medical Association, 282*(22), 2131–2135.

Walter, R. B., Brasky, T. M., Buckley, S. A., Potter, J. D., & White, E. (2013). Height as an explanatory factor for sex differences in human cancer. *Journal of the National Cancer Institute, 105*(12), 860–868. https://doi.org/10.1093/jnci/djt102

Wang, Z., Wang, D., & Wang, Y. (2017). Cigarette smoking and adipose tissue: The emerging role in progression of atherosclerosis. *Mediators of Inflammation, 2017*. https://doi.org/10.1155/2017/3102737

West, H. (Jack), & Jin, J. O. (2015). Performance status in patients with cancer. *JAMA Oncology, 1*(7), 998–998. https://doi.org/10.1001/jamaoncol.2015.3113

Winkleby, M. A., Jatulis, D. E., Frank, E., & Fortmann, S. P. (1992). Socioeconomic status and health: How education, income, and occupation contribute to risk factors for cardiovascular disease. *American Journal of Public Health*, 82(6), 816–820.

World Cancer Research Fund/American Institute for Cancer Research. Diet, nutrition, physical activity and lung cancer. In: Continuous Update Project Expert Report 2018 [Internet]. [cited 7 March 2019]. Available from: dietandcancerreport.org, https://www.wcrf.org/dietandcancer

World Cancer Research Fund/American Institute for Cancer Research. Continuous Update Project Export

Report 2018. Meat, fish and dairy products and the risk of cancer. Available at dietandcancerreport.org

World Health Organization. (n.d.). *Risk factors.* Retrieved August 9, 2018, from http://www.who.int/topics/risk_factors/en/

Howard, J. (2015). Minimum latency & types or categories of cancer. *Policies and Procedures – World Trade Center Health Program.* Retrieved June 22, 2019, from https://www.cdc.gov/wtc/policies.html

Willett WC, Trichopoulos D. (1996). Nutrition and cancer: A summary of the evidence. *Cancer Causes Control, 7*(1):178–80.

Wynder, Ernest L., & Graham, E. A. (1950). Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. *Bulletin of the World Health Organization, 83*(2), 146–153.

Wynder, Ernst L, & Muscat, J. E. (1995). The changing epidemiology of smoking and lung cancer histology. *Environmental Health Perspectives, 103*, 6.

Yang, W.S., Va, P., Wong, M.Y., Zhang, H.L., & Xiang, Y.B. (2011). Soy intake is associated with lower lung cancer risk: Results from a meta-analysis of epidemiologic studies. *The American Journal of Clinical Nutrition, 94*(6), 1575–1583. https://doi.org/10.3945/ajcn.111.020966

Yang, Y., Dong, J., Sun, K., Zhao, L., Wang, L., & Jiao, Y. (2012). Obesity and incidence of lung cancer: A meta-analysis. *International Journal of Cancer, 132*(5), 1162–1169. https://doi.org/10.1002/ijc.27719

Yokota, J., Shiraishi, K., & Kohno, T. (2010). Genetic basis for susceptibility to lung cancer: Recent progress and future directions. In G. F. Vande Woude & G. Klein (Eds.), *Advances in Cancer Research* (pp. 51–72). Location: Academic Press. https://doi.org/10.1016/B978-0-12-380890-5.00002-8

Yong, L.C., Brown, C., Schatzkin, A., Dresser, C., Slesinski, M. J., Cox, C., & Taylor, P. (1997). Intake of

vitamins E, C, and A and risk of lung cancer the NHANES I Epidemiologic followup Study. *American Journal of Epidemiology, 146*(3), 231–243. https://doi.org/10.1093/oxfordjournals.aje.a009258

Young, R. P., Duan, F., Chiles, C., Hopkins, R. J., Gamble, G. D., Greco, E. M., … Aberle, D. (2015). Airflow limitation and histology shift in the national lung screening trial. The NLST-ACRIN cohort substudy. *American Journal of Respiratory and Critical Care Medicine, 192*(9), 1060–1067. https://doi.org/10.1164/rccm.201505-0894OC

Young, R. P., & Hopkins, R. J. (2010). Link between COPD and lung cancer. *Respiratory Medicine, 104*(5), 758–759. https://doi.org/10.1016/j.rmed.2009.11.025

Yu, Z. M., DeClercq, V., Cui, Y., Forbes, C., Grandy, S., Keats, M., … Dummer, T. J. B. (2018). Fruit and vegetable intake and body adiposity among populations in Eastern Canada: The Atlantic partnership for tomorrow's health study. *BMJ Open, 8*(4). https://doi.org/10.1136/bmjopen-2017-018060

Zhou, Y., & Rui, L. (2013). Leptin signaling and leptin resistance. *Frontiers of medicine*, 7(2), 207-22. doi: 10.1007/s11684-013-0263-5

Zhou, W., & Christiani, D. C. (2011). East meets West: Ethnic differences in epidemiology and clinical behaviors of lung cancer between East Asians and Caucasians. *Chinese Journal of Cancer, 30*(5), 287–292. https://doi.org/10.5732/cjc.011.10106

SUPPLEMENTAL MATERIAL

Table S1. Characteristic of Overall Participants by Lung Cancer Status and Univariate Logistic Associations with Lung Cancer

Explanatory Variables	Controls (n = 1919)	Cases (n = 931)	*P*	Univariate Odds Ratio (95% CI; *P*)
Sociodemographic				
Age, year, mean (SD)	64.10 (5.04)	64.11 (5.01)	$P_t = 0.933$	1.161 (0.997–1.352; 0.055)
Sex, number (%)			$P_c = 0.531$	Controls matched to cases
Female	780 (40.65)	367 (39.42)		
Male	1139 (59.35)	564 (60.58)		
Race/ethnicity, number (%)			$P_c = 0.871$	Controls matched to cases
Non-Hispanic White	1734 (90.36)	839 (90.12)		
Non-Hispanic Black	112 (5.84)	57 (6.12)		
Hispanic	26 (1.35)	12 (1.29)		
Asian	35 (1.82)	17 (1.83)		
Pacific Islander	6 (0.31)	5 (0.54)		
American Indian/Alaskan Native	6 (0.31)	1 (0.11)		
Education, number (%)			$P_c < 0.001$	
Less than 8 years	18 (0.94)	13 (1.40)		1.310 (0.621–2.762; 0.479)
8 to 11 years	153 (7.99)	95 (10.20)		1.142 (0.836–1.561; 0.403)
12 years or completed high school	435 (22.72)	231 (24.81)		Referent group
Post high school training other than college	264 (13.79)	136 (14.61)		0.963 (0.738–1.258; 0.784)
Some college	440 (22.98)	234 (25.13)		1.016 (0.807–1.280; 0.892)
College graduate	295 (15.40)	128 (13.75)		0.841 (0.645–1.098; 0.204)
Postgraduate degree	310 (16.19)	94 (10.10)		0.580 (0.434–0.776; <0.001)
				<0.001(overall)
Medical history				
BMI at age 50, kg/m^2, mean (SD)	25.26 (3.83)	25.28 (3.50)	$P_t = 0.929$	0.965 (0.942–0.989; 0.005)
Self-reported COPD, number (%)			$P_c < 0.001$	
No	1757 (91.56)	788 (84.64)		
Yes	162 (8.44)	143 (15.36)		$1.850_{\text{yes vs no}}$ (1.447–2.365; <0.001)

Abbreviations: BMI, body mass index (weight kg/height meter^2); CI, confidence interval; COPD, chronic obstructive pulmonary disease; CP, C-peptide; hsCRP, high-sensitivity C-reactive protein; IQR, interquartile range; LN, natural log transformed; *P*, P-value; P_c: P-value from Pearson's chi-squared; P_e, P-value from Fisher's Exact test; P_n, P-value from non-parametric test for trend across ordered groups (an extension of the Wilcoxon rank-sum test); P_t, P-value from t-test with unequal variances; SD, standard deviation.

(continued on the following page)

Explanatory Variables	Controls (n = 1919)	Cases (n = 931)	*P*	Univariate Odds Ratio (95% CI; *P*)
Dietary factors				
Food energy from diet, kcal/day, median (IQR)	1978.44 (1058.58)	1943.85 (1041.71)	$P_n = 0.542$	
LN Food energy from diet, mean (SD)	7.58 (0.43)	7.58 (0.42)	$P_t = 0.844$	0.982 (0.796–1.211; 0.866)
Fruits&Vegetables, daily frequency, median (IQR)	6.01 (4.12)	5.42 (3.93)	$P_n < 0.001$	
LN Fruits&Vegetables, mean (SD)	1.91 (1.16)	1.63 (1.22)	$P_t < 0.001$	0.816 (0.756–0.882; <0.001)
Red meat, g/day, median (IQR)	24.58 (35.06)	28.28 (38.27)	$P_n = 0.001$	
LN Red meat, mean (SD)	3.16 (0.97)	3.31 (0.95)	$P_t < 0.001$	1.149 (1.042–1.267; 0.005)
Supplemental beta-carotene, 1000 mcg/day, mean (SD)	0.52 (1.00)	0.41 (0.83)	$P_t = 0.006$	0.864 (0.782–0.954; 0.004)
Vitamin C from diet, mg/day, median (IQR)	150.69 (104.61)	132.75 (105.59)	$P_n < 0.001$	
LN Vitamin C from diet, mean (SD)	4.96 (0.59)	4.85 (0.60)	$P_t < 0.001$	0.756 (0.654–0.875; <0.001)
Isoflavone, mg/day, median (IQR)	0.40 (0.48)	0.36 (0.43)	$P_n = 0.039$	
LN Isoflavone, median (SD)	-0.98 (1.03)	-1.05 (0.96)	$P_t = 0.088$	0.919 (0.838–1.009; 0.077)
Smoking Exposures				
Smoking status, number of current-smokers (%)	772 (40.23)	393 (42.21)	$P_c = 0.005$	Controls matched to cases
Cigarettes per day, mean (SD)	20.71 (14.72)	26.86 (16.30)	$P_t < 0.001$	1.033 (1.026–1.040; <0.001)
Smoking duration, year, mean (SD)	28.32 (17.22)	39.91 (14.80)	$P_t < 0.001$	1.077 (1.065–1.088; <0.001)
Quit-time in former-smokers, year, mean (SD)	23.26 (12.51)	13.16 (10.45)	$P_t < 0.001$	0.929 (0.917–0.940; <0.001)
Metabolic markers				
CP (pmol/L), median (IQR)	752.26 (670.17)	811.44 (756.89)	$P_n = 0.025$	
LN CP, mean (SD)	6.65 (0.64)	6.70 (0.64)	$P_t = 0.060$	1.127 (0.993–1.281; 0.065)
hsCRP (mg/L), median (IQR)	8.40 (14.90)	11.05 (20.45)	$P_n < 0.001$	
LN hsCRP, mean (SD)	2.32 (1.01)	2.56 (1.04)	$P_t < 0.001$	1.292 (1.186–1.407; <0.001)
Leptin (ng/mL), median (IQR)	5.30 (7.50)	5.20 (7.40)	$P_n = 0.174$	
LN Leptin, mean (SD)	1.89 (0.78)	1.84 (0.79)	$P_t = 0.134$	0.932 (0.828–1.050; 0.247)

Abbreviations: BMI, body mass index (weight kg/height meter^2); CI, confidence interval; COPD, chronic obstructive pulmonary disease; CP, C-peptide; hsCRP, high-sensitivity C-reactive protein; IQR, interquartile range; LN, natural log transformed; *P*, P-value; P_c: P-value from Pearson's chi-squared; P_e, P-value from Fisher's Exact test; P_n, P-value from non-parametric test for trend across ordered groups (an extension of the Wilcoxon rank-sum test); P_t, P-value from t-test with unequal variances; SD, standard deviation.

Table S2. Characteristic of Never-smokers by Lung Cancer Status and Univariate Logistic Associations with Lung Cancer

Explanatory Variables	Controls (n = 234)	Cases (n = 76)	*P*	Univariate Odds Ratio (95% CI; *P*)
Sociodemographic				
Age, year, mean (SD)	65.21 (5.30)	65.66 (5.29)	$P_t = 0.526$	1.070 (0.632–1.812; 0.800)
Sex, number (%)			$P_c = 0.621$	Controls matched to cases
Female	149 (63.68)	46 (60.53)		
Male	85 (36.32)	30 (39.47)		
Race/ethnicity, number (%)			$P_e = 0.957$	Controls matched to cases
Non-Hispanic White	220 (94.02)	71 (93.42)		
Non-Hispanic Black	3 (1.28)	1 (1.32)		
Hispanic	5 (2.14)	2 (2.63)		
Asian	6 (2.56)	2 (2.63)		
Pacific Islander	0 (0.00)	0 (0.00)		
American Indian/Alaskan Native	0 (0.00)	0 (0.00)		
Education, number (%)			$P_e = 0.053$	
Less than 8 years	0 (0.00)	2 (2.63)		N/A
8 to 11 years	15 (6.41)	4 (5.26)		1.316 (0.333–5.205; 0.696)
12 years or completed high school	54 (23.08)	9 (11.84)		Referent group
Post high school training other than college	30 (12.82)	14 (18.42)		2.445 (0.942–6.350; 0.066)
Some college	55 (23.50)	14 (18.42)		1.498 (0.585–3.835; 0.399)
College graduate	37 (15.81)	15 (19.74)		2.529 (0.938–6.819; 0.067)
Postgraduate degree	43 (18.38)	18 (23.68)		2.534 (0.965–6.655; 0.059)
				0.206 (overall)
Medical history				
BMI at age 50, kg/m^2, mean (SD)	25.32 (3.87)	25.05 (3.72)	$P_t = 0.580$	0.977 (0.906–1.053; 0.537)
Self-reported COPD, number (%)			$P_e = 0.462$	
No	226 (96.58)	75 (98.68)		
Yes	8 (3.42)	1 (1.32)		$0.383_{\text{yes vs no}}$ (0.047–3.123; 0.370)

Abbreviations: BMI, body mass index (weight kg/height meter^2); CI, confidence interval; COPD, chronic obstructive pulmonary disease; CP, C-peptide; hsCRP, high-sensitivity C-reactive protein; IQR, interquartile range; LN, natural log transformed; *P*, P-value; P_c: P-value from Pearson's chi-squared; P_e, P-value from Fisher's Exact test; P_n, P-value from non-parametric test for trend across ordered groups (an extension of the Wilcoxon rank-sum test); P_t, P-value from t-test with unequal variances; SD, standard deviation.

(continued on the following page)

Explanatory Variables	Controls (n = 234)	Cases (n = 76)	*P*	Univariate Odds Ratio (95% CI; *P*)
Dietary factors				
Food energy from diet, kcal/day, median (IQR)	1707.13 (855.68)	1981.11 (918.90)	$P_n = 0.035$	
LN Food energy from diet, mean (SD)	7.43 (0.44)	7.55 (0.36)	$P_t = 0.022$	1.764 (0.913–3.408; 0.091)
Fruits&Vegetables, daily frequency, median (IQR)	6.83 (4.05)	7.16 (3.71)	$P_n = 0.147$	
LN Fruits&Vegetables, mean (SD)	2.23 (1.01)	2.42 (0.82)	$P_t = 0.108$	1.233 (0.921–1.650; 0.159)
Red meat, g/day, median (IQR)	16.30 (22.25)	15.89 (26.31)	$P_n = 0.833$	
LN Red meat, mean (SD)	2.79 (0.94)	2.88 (0.89)	$P_t = 0.467$	1.069 (0.787–1.453; 0.669)
Supplemental beta-carotene, 1000 mcg/day, mean (SD)	0.58 (1.07)	0.40 (0.56)	$P_t = 0.066$	0.736 (0.502–1.080; 0.118)
Vitamin C from diet, mg/day, median (IQR)	154.36 (103.62)	169.08 (93.96)	$P_n = 0.104$	
LN Vitamin C from diet, mean (SD)	5.02 (0.54)	5.14 (0.49)	$P_t = 0.091$	1.501 (0.891–2.529; 0.127)
Isoflavone, mg/day, median (IQR)	0.37 (0.50)	0.41 (0.44)	$P_n = 0.553$	
LN Isoflavone, median (SD)	-1.00 (1.08)	-0.94 (0.92)	$P_t = 0.622$	1.117 (0.836–1.492; 0.455)
Smoking Exposures				
Smoking status, number of never-smokers (%)	234 (100.00)	76 (100.00)	N/A	Controls matched to cases
Metabolic markers				
CP (pmol/L), median (IQR)	789.14 (686.91)	706.50 (517.28)	$P_n = 0.457$	
LN CP, mean (SD)	6.65 (0.63)	6.59 (0.59)	$P_t = 0.497$	0.870 (0.571–1.325; 0.515)
hsCRP (mg/L), median (IQR)	7.95 (14.80)	7.70 (14.20)	$P_n = 0.769$	
LN hsCRP, mean (SD)	2.28 (1.03)	2.24 (1.02)	$P_t = 0.741$	0.993 (0.754–1.308; 0.959)
Leptin (ng/mL), median (IQR)	7.70 (13.00)	6.20 (8.20)	$P_n = 0.246$	
LN Leptin, mean (SD)	2.15 (0.83)	2.04 (0.72)	$P_t = 0.283$	0.872 (0.586–1.296; 0.497)

Abbreviations: BMI, body mass index (weight kg/height meter^2); CI, confidence interval; COPD, chronic obstructive pulmonary disease; CP, C-peptide; hsCRP, high-sensitivity C-reactive protein; IQR, interquartile range; LN, natural log transformed; *P*, P-value; P_c: P-value from Pearson's chi-squared; P_e, P-value from Fisher's Exact test; P_n, P-value from non-parametric test for trend across ordered groups (an extension of the Wilcoxon rank-sum test); P_t, P-value from t-test with unequal variances; SD, standard deviation.

Table S3. Multivariable Conditional Logistic Regression for Dietary Factors Associated with Lung Cancer in 2062 Ever-Smokers

Explanatory Variables	Beta Coefficient (95% CI)	Odds Ratio (95% CI)	*P*
Education, per 1 level increase	-0.062 (-0.127–0.003)	0.940 (0.881–1.003)	0.061
LN Fruits&Vegetables, per 1 ln(daily frequency) increase	-0.127 (-0.219– -0.036)	0.881 (0.803–0.965)	0.007
Supplemental beta-carotene, per 1000 mcg/day increase	-0.143 (-.257– -0.029)	0.867 (0.773–0.971)	0.014
Smoking intensity, per 1 cigarette/day increase	-1.731 (-2.206– -1.257)†	N/A	<0.001
Smoking duration, per 1-year increase	0.066 (.0537–0.0784)	1.068 (1.055–1.082)	<0.001
n; AIC, BIC	n = 2062; 1201.649, 1229.806		
P of overall model, pseudo-R^2	<0.001, 0.203		

† The association between smoking intensity (average number of cigarettes smoked per day) and lung cancer was non-linear, and the variable was transformed accordingly. As a result, the effect estimate could not be interpreted directly.
The transformation formula: (number of cigarettes per day/10)^-1

Table S4. Multivariable Conditional Logistic Regression for Body Mass Index (BMI) Associated with Lung Cancer in 2340 Ever-Smokers

Explanatory Variables	Beta Coefficient (95% CI)	Odds Ratio (95% CI)	*P*
Age, per 1-year increase	0.171 (-0.005– -0.348)	1.187 (0.995–1.416)	0.057
Education, per 1 level increase	-0.077 (-.1386– -0.016)	0.926 (0.871–0.984)	0.014
BMI at age 50, per 1 kg/m^2 increase	-0.024 (-0.054–0.006)	0.976 (0.947–1.006)	0.111
Self-reported COPD, yes versus no	0.242 (-0.043–0.528)	1.274 (0.958–1.695)	0.096
Smoking intensity, per 1 cigarette/day increase	-1.712 (-2.148– -1.275)†	N/A†	<0.001
Smoking duration, per 1-year increase	0.048 (0.027–0.069)	1.049 (1.028–1.072)	<0.001
Quit-time, per 1-year increase	-0.019 (-0.043–0.005)	0.981 (0.958–1.005)	0.122
Interaction between BMI and age	0.007 (0.001–0.013)	1.007 (1.001–1.013)	0.024
n; AIC, BIC	n = 2340; 1399.280, 1445.343		
P of overall model, pseudo-R^2	<0.001, 0.188		

† A non-linear association between smoking intensity and lung cancer odds could not be interpreted directly.
The transformation formula: (number of cigarettes per day/10)^-1

Table S5. Multivariable Linear Regression with Dietary Factors Predicting Body Mass Index (BMI) in 2261 Ever-Smokers

Explanatory Variables	Beta Coefficient (95% CI)	*P*
Age, per 1-year increase	-0.080 (-0.114– -0.045)	<0.001
Education, per 1 level increase	-0.053 (-0.140–0.035)	0.240
LN Food energy from diet, per 1 ln(kcal/day) increase	0.600 (0.137–1.064)	0.011
LN Fruits&Vegetables, per 1 ln(daily frequency) increase	-0.407 (-0.660– -0.155)	0.002
LN Red meat, per 1 ln(g/day) increase	0.388 (0.215–0.561)	<0.001
Supplemental beta-carotene, per 1000 mcg/day increase	0.00006 (0.00003–0.00010)†	<0.001
LN Vitamin C from diet, per 1 ln(mg/day) increase	0.875 (0.350–1.399)	0.001
LN Isoflavone, per 1 ln(mg/day) increase	-0.243 (-0.405– -0.082)	0.003
Smoking intensity, per 1 cigarette/day increase	0.022 (0.011–0.032)	<0.001
Smoking duration, per 1-year increase	-0.012 (-0.031–0.007)	0.210
Quit-time, per 1-year increase	0.238 (0.096–0.380)‡	0.001
Model constant	-12.766 (-16.172– -9.360)	<0.001
n; AIC, BIC	n = 2261; 11827.700, 11896.380	
P of overall model, R^2	<0.001, 0.093	

† According to the multivariable fractional polynomial (MFP) model in Stata, supplemental beta-carotene was transformed to a power of -2. Therefore, no direct interpretation for the coefficient estimate was attainable.
The transformation formula: (supplemental beta-carotene+ 0.0107249990105629))^-2
‡ Quit-time was transformed at the suggestion of MFP. The transformation formula: ln [(quit-time+0.5)/10]

Table S6. Multivariable Conditional Logistic Regression for Dietary Factors and Body Mass Index (BMI) Associated with Lung Cancer in 2027 Ever-Smokers

Explanatory Variables	Beta Coefficient (95% CI)	Odds Ratio (95% CI)	*P*
Education, per 1 level increase	-0.075 (-0.141– -0.009)	0.928 (0.869–0.991)	0.026
BMI at age 50, per 1 kg/m^2 increase	-0.040 (-0.073– -0.007)	0.961 (0.930–0.993)	0.017
LN Fruits&Vegetables, per 1 ln(daily frequency) increase	-0.129 (-0.222– -0.036)	0.879 (0.801–0.964)	0.006
Supplemental beta-carotene, per 1000 mcg/day increase	-0.150 (-0.265– -0.035)	0.861 (0.767–0.965)	0.010
Smoking intensity, per 1 cigarette/day increase	-1.719 (-2.197– -1.240)†	N/A†	<0.001
Smoking duration, per 1-year increase	0.065 (0.052–0.077)	1.067 (1.053–1.080)	<0.001
n; AIC, BIC	n = 2027; 1178.660, 1212.346		
P of overall model, pseudo-R^2	<0.001, 0.205		

† A non-linear association between smoking intensity and lung cancer odds could not be interpreted directly.
The transformation formula: (number of cigarettes per day/10)^-1

Table S7a. Multivariable Linear Regression with Dietary Factors Predicting C-peptide (CP) in 2283 Ever-Smokers

Explanatory Variables	Beta Coefficient (95% CI)	*P*
Age, per 1-year increase	0.008 (0.003–0.014)	0.004
Sex, female versus male	-0.087 (-0.144– -0.029)	0.003
LN Fruits&Vegetables, per 1 ln(daily frequency) increase	0.030 (-0.001–0.060)	0.057
LN Red meat, per 1 ln(g/day) increase	0.053 (0.024–0.083)	<0.001
Smoking intensity, per 1 cigarette/day increase	0.003 (0.001–0.005)	0.012
Interaction between sex and LN Fruits&Vegetables	-0.057 (-0.104– -0.011)	0.016
Model constant	-0.533 (-0.649– -0.417)	<0.001
n; AIC, BIC	n = 2283; 4371.563, 4411.696	
P of overall model, R^2	<0.001, 0.024	

Table S7b. Multivariable Linear Regression with Dietary Factors Predicting High-sensitivity C-reactive Protein (hsCRP) in 2281 Ever-Smokers

Explanatory Variables	Beta Coefficient (95% CI)	*P*
Age, per 1-year increase	-0.020 (-0.048–0.007)	0.152
Sex, female versus male	0.653 (0.338–0.967)	<0.001
Education, per 1 level increase	-0.022 (-0.049–0.005)	0.114
LN Fruits&Vegetables, per 1 ln(daily frequency) increase	-0.049 (-0.084– -0.014)	0.007
LN Red meat, per 1 ln(g/day) increase	0.113 (0.060–0.167)	<0.001
Self-reported COPD, yes versus no	0.304 (0.167–0.441)	<0.001
Smoking intensity, per 1 cigarette/day increase	0.004 (0.001–0.007)	0.004
Smoking duration, per 1-year increase	0.010 (0.007–0.013)	<0.001
Interaction between sex and LN Red meat	-0.111 (-0.209– -0.013)	0.026
Interaction between age and LN Red meat	0.007 (-0.001–0.015)	0.075
Model constant	-0.277 (-0.488– -0.066)	0.010
n; AIC, BIC	n = 2281; 6436.229, 6499.285	
P of overall model, R^2	<0.001, 0.069	

Table S7c. Multivariable Linear Regression with Dietary Factors Predicting Leptin in 2274 Ever-Smokers

Explanatory Variables	Beta Coefficient (95% CI)	*P*
Sex, female versus male	0.787 (0.718–0.855)	<0.001
Education, per 1 level increase	0.012 (-0.007–0.030)	0.217
LN Fruits&Vegetables, per 1 ln(daily frequency) increase	-0.051 (-0.102– -0.001)	0.047
LN Red meat, per 1 ln(g/day) increase	0.078 (0.044–0.111)	<0.001
Supplemental beta-carotene, per 1000 mcg/day increase	0.00001 (0.000007–0.00002)†	<0.001
LN Vitamin C from diet, per 1 ln(mg/day) increase	0.101 (0.008–0.193)	0.033
LN Isoflavone, per 1 ln(mg/day) increase	-0.045 (-0.077– -0.013)	0.006
Smoking intensity, per 1 cigarette/day increase	0.003 (0.001–0.005)	0.008
Quit-time, per 1-year increase	-0.0007 (-0.0009– -0.0006)‡	0.001
Model constant	-1.862 (-2.374– -1.351)	<0.001
n; AIC, BIC	n = 2274; 4704.466, 4761.759	
P of overall model, R^2	<0.001, 0.226	

† The transformation formula: (supplemental beta-carotene+0.0107249990105629)^-2

‡ The transformation formula: [(quit-time+0.5)/10]^-2

Table S8a. Multivariable Linear Regression with Body Mass Index (BMI) Predicting C-peptide (CP) in 2461 Ever-Smokers

Explanatory Variables	Beta Coefficient (95% CI)	*P*
Age, per 1-year increase	0.009 (0.004–0.014)	<0.001
BMI at age 50, per 1 kg/m^2 increase	0.425 (0.334–0.516)†	<0.001
BMI at age 50, per 1 kg/m^2 increase	-0.382 (-0.496– -0.267)‡	<0.001
Smoking intensity, per 1 cigarette/day increase	0.003 (0.001–0.004)	0.006
Smoking duration, per 1-year increase	0.001 (-0.001–0.003)	0.142
Model constant	-0.796 (-0.888– -0.705)	<0.001
n; AIC, BIC	n = 2461; 4685.637, 4720.487	
P of overall model, R^2	<0.001, 0.054	

† The transformation formula: [(centred BMI+11.36660194396973)/10]^2

‡ The transformation formula: [(centred BMI+11.36660194396973)/10]^2*ln[(centred BMI+11.36660194396973)/10]

Table S8b. Multivariable Linear Regression with Body Mass Index (BMI) Predicting High-sensitivity C-reactive Protein (hsCRP) in 2457 Ever-Smokers

Explanatory Variables	Beta Coefficient (95% CI)	*P*
Sex, female versus male	0.368 (0.261–0.474)	<0.001
Non-Hispanic Black versus White	0.189 (0.011–0.367)	0.038
Hispanic versus White	0.225 (-0.444–0.894)	0.510
Asian versus White	-0.629 (-0.927– -0.331)	<0.001
Pacific Islander versus White	-0.555 (-1.084– -0.025)	0.040
American Indian/Alaskan Native versus White	0.109 (-0.295–0.513)	0.598
Education, per 1 level increase	-0.029 (-0.055– -0.003)	0.027
Self-reported COPD, yes versus no	0.257 (0.128–0.387)	<0.001
BMI at age 50, per 1 kg/m^2 increase	0.021 (0.006–0.035)	0.005
Smoking intensity, per 1 cigarette/day increase	0.005 (0.002–0.008)	<0.001
Smoking duration, per 1-year increase	0.012 (0.009–0.015)	<0.001
Interaction between sex and BMI	0.035 (0.010–0.060)	0.005
Model constant	-0.226 (-0.395– -0.056)	0.009
n; AIC, BIC	n = 2457; 6900.960, 6976.447	
P of overall model, R^2	<0.001, 0.086	

Table S8c. Multivariable Linear Regression with Body Mass Index (BMI) Predicting Leptin in 2455 Ever-Smokers

Explanatory Variables	Beta Coefficient (95% CI)	*P*
Age, per 1-year increase	0.007 (0.001–0.012)	0.018
Sex, female versus male	0.918 (0.865–0.972)	<0.001
Non-Hispanic Black versus White	0.197 (0.093–0.302)	<0.001
Hispanic versus White	-0.015 (-0.238–0.208)	0.893
Asian versus White	-0.171 (-0.333– -0.009)	0.039
Pacific Islander versus White	0.048 (-0.243–0.340)	0.746
American Indian/Alaskan Native versus White	-0.155 (-0.742–0.431)	0.603
Education, per 1 level increase	0.014 (-0.002–0.030)	0.086
BMI at age 50, per 1 kg/m^2 increase	1.964 (1.686–2.242)†	<0.001
BMI at age 50, per 1 kg/m^2 increase	-0.448 (-0.561– -0.335)‡	<0.001
Smoking intensity, per 1 cigarette/day increase	0.002 (0.001–0.004)	0.014
Smoking duration, per 1-year increase	0.004 (0.002–0.007)	0.001
Quit-time, per 1-year increase	-0.0006 (-0.0008– -0.0004)§	0.001
Model constant	-2.616 (-2.814– -2.419)	<0.001
n; AIC, BIC	n = 2455; 4506.633, 4587.915	
P of overall model, R^2	<0.001, 0.397	

† The transformation formula: (centred BMI +11.36660194396973)/10

‡ The transformation formula: [(centred BMI+11.36660194396973)/10]^2

§ The transformation formula: [(quit-time+0.5)/10]^-2

Table S9. Comparison of Multivariable Conditional Logistic Regressions for Metabolic Markers, Body Mass Index (BMI) and Metabolic Markers, and Dietary Factors and Metabolic Markers Associated with Lung Cancer in Ever-Smokers

Explanatory Variables	Model 1 Odds Ratio (95% CI), *P*	Model 2 Odds Ratio (95% CI), *P*	Model 3 Odds Ratio (95% CI), *P*
Education, per 1 level increase	0.929 (0.875–0.987), 0.018	0.921 (0.866–0.980), 0.009	0.941 (0.881–1.004), 0.068
Smoking intensity, per 1 cigarette/day increase	N/A†, <0.001	N/A†, <0.001	N/A†, <0.001
Smoking duration, per 1-year increase	1.066 (1.054–1.078), <0.001	1.065 (1.053–1.077), <0.001	1.068 (1.055–1.082), <0.001
LN Fruits&Vegetables, per 1 ln(daily frequency) increase	-	-	0.893 (0.813–0.980), 0.017
Supplemental beta-carotene, per 1000 mcg/day increase	-	-	0.859 (0.764–0.967), 0.012
BMI at age 50, per 1 kg/m2 increase	-	0.973 (0.941–1.007), 0.118	-
LN CP, per 1 ln(pmol/L) increase	1.205 (1.023–1.420), 0.025	1.229 (1.043–1.448), 0.014	1.254 (1.050–1.495), 0.012
LN hsCRP, per 1 ln(mg/L) increase	1.233 (1.110–1.371), <0.001	1.228 (1.104–1.366), <0.001	1.212 (1.080–1.359), 0.001
LN Leptin, per 1 ln(ng/mL) increase	0.746 (0.638–0.872), <0.001	0.806 (0.677–0.961), 0.016	0.704 (0.592–0.837), <0.001
n; AIC, BIC	n = 2348; 1386.992, 1421.560	n = 2308; 1363.730, 1403.939	n = 2037; 1165.690, 1210.644
P of overall model, pseudo-R^2	<0.001, 0.196	<0.001, 0.196	<0.001, 0.221

† A non-linear association between smoking intensity and lung cancer odds could not be interpreted directly.
The transformation formula: (number of cigarettes per day/10)^-1.